METHODS TO ENGINEER MAMMALIAN-TYPE CARBOHYDRATE STRUCTURES

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to U. S. provisional application Ser. No. 60/344,169, Dec. 27, 2001, which is incorporated by reference herein in its entirety.

FIELD OF THE INVENTION

[0002] The present invention generally relates to modifying the glycosylation structures of recombinant proteins expressed in fungi or other lower eukaryotes, to more closely resemble the glycosylation of proteins of higher mammals, in particular humans.

BACKGROUND OF THE INVENTION

[0003] After DNA is transcribed and translated into a protein, further post translational processing involves the attachment of sugar residues, a process known as glycosylation. Different organisms produce different glycosylation enzymes (glycosyltransferases and glycosidases), and have different substrates (nucleotide sugars) available, so that the glycosylation patterns as well as composition of the individual oligosaccharides, even of one and the same protein, will be different depending on the host system in which the particular protein is being expressed. Bacteria typically do not glycosylate proteins, and if so only in a very unspecific manner (Moens, 1997). Lower eukaryotes such as filamentous fungi and yeast add

primarily mannose and mannosylphosphate sugars, whereas insect cells such as Sf9 cells glycosylate proteins in yet another way. See for example (Bretthauer, 1999; Martinet, 1998; Weikert, 1999; Malissard, 2000; Jarvis, 1998; and Takeuchi, 1997).

[0004] Synthesis of a mammalian-type oligosaccharide structure consists of a series of reactions in the course of which sugar-residues are added and removed while the protein moves along the secretory pathway in the host organism. The enzymes which reside along the glycosylation pathway of the host organism or cell determine what the resulting glycosylation patterns of secreted proteins. Unfortunately, the resulting glycosylation pattern of proteins expressed in lower eukaryotic host cells differs substantially from the glycosylation found in higher eukaryotes such as humans and other mammals (Bretthauer, 1999). Moreover, the vastly different glycosylation pattern has, in some cases, been shown to increase the immunogenicity of these proteins in humans and reduce their half-life (Takeuchi, 1997). It would be desirable to produce human-like glycoproteins in non-human host cells, especially lower eukaryotic cells.

[0005] The early steps of human glycosylation can be divided into at least two different phases: (i) lipid-linked $Glc_3Man_9GlcNAc_2$ oligosaccharides are assembled by a sequential set of reactions at the membrane of the endoplasmic reticulum (ER) and (ii) the transfer of this oligosaccharide from the lipid anchor dolichyl pyrophosphate onto *de novo* synthesized protein. The site of the specific transfer is defined by an asparagine (Asn) residue in the sequence Asn-Xaa-Ser/Thr (see **Fig. 1**), where Xaa can be any amino acid except proline (Gavel, 1990). Further processing by glucosidases and mannosidases occurs in the ER before the nascent glycoprotein is transferred to the early Golgi apparatus, where additional mannose residues are removed by Golgi specific alpha (α)-1,2-mannosidases. Processing continues as the protein proceeds through the Golgi. In the medial Golgi, a number of modifying enzymes, including N-acetylglucosaminyltransferases (GnT I, GnT II, GnT III, GnT IV GnT V GnT VI), mannosidase II and fucosyltransferases, add and remove specific sugar residues (see, e.g., **Figs. 2 and 3**). Finally, in the trans-Golgi, galactosyltranferases and sialyltransferases produce a glycoprotein structure that is released from the Golgi. It is this structure,

characterized by bi-, tri- and tetra-antennary structures, containing galactose, fucose, N-acetylglucosamine and a high degree of terminal sialic acid, that gives glycoproteins their human characteristics.

[0006] In nearly all eukaryotes, glycoproteins are derived from the common core oligosaccharide precursor $Glc_3Man_9GlcNAc_2$-PP-Dol, where PP-Dol stands for dolichol-pyrophosphate (**Fig. 1**). Within the endoplasmic reticulum, synthesis and processing of dolichol pyrophosphate bound oligosaccharides are identical between all known eukaryotes. However, further processing of the core oligosaccharide by yeast, once it has been transferred to a peptide leaving the ER and entering the Golgi, differs significantly from humans as it moves along the secretory pathway and involves the addition of several mannose sugars.

[0007] In yeast, these steps are catalyzed by Golgi residing mannosyltransferases, like Och1p, Mnt1p and Mnn1p, which sequentially add mannose sugars to the core oligosaccharide. The resulting structure is undesirable for the production of humanoid proteins and it is thus desirable to reduce or eliminate mannosyltransferase activity. Mutants of *S. cerevisiae*, deficient in mannosyltransferase activity (for example *och1* or *mnn9* mutants) have been shown to be non-lethal and display a reduced mannose content in the oligosacharide of yeast glycoproteins. Other oligosacharide processing enzymes, such as mannosylphophate transferase may also have to be eliminated depending on the host's particular endogenous glycosylation pattern.

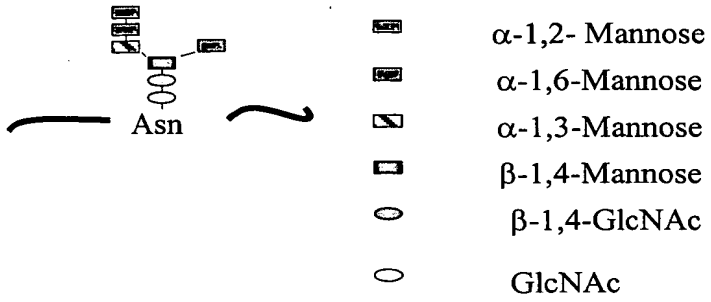
Lipid-Linked Oligosaccharide Precursors

[0008] Of particular interest for this invention are the early steps of N-glycosylation (**Figs. 1 and 2**). The study of *alg* (asparagine-linked glycosylation) mutants defective in the biosynthesis of the $Glc_3Man_9GlcNAc_2$-PP-Dol has helped to elucidate the initial steps of N-glycosylation.

[0009] The *ALG3* gene of *S.cerevisiae* has been succesfully cloned and knocked out by deletion (Aebi, 1996). *ALG3* has been shown to encode the enzyme Dol-P-Man:$Man_5GlcNAc_2$-PP-Dol Mannosyltransferase, which is involved in the first Dol-P-Man dependent mannosylation step from $Man_5GlcNAc_2$-PP-Dol to $Man_6GlcNAc_2$-PP-Dol at the luminal side of the ER (Sharma, 2001) (**Figs 1 and**

2). *S.cerevisiae* cells harboring a leaky *alg3-1* mutation accumulate $Man_5GlcNAc_2$-PP-Dol (structure I) (Huffaker, 1983).

Structure I: $Man_5GlcNAc_2$



$Man_5GlcNAc_2$ (Structure I) and $Man_8GlcNAc_2$ accumulate in total cell mannoprotein of an *och1 mnn1 alg3* mutant(Nakanishi-Shindo, 1993). This *S.cerevisiae och1, mnn1, alg3* mutant was shown to be viable, but temperature-sensitive, and to lack α–1,6 polymannose outer chains.

[0010] In another study, secretory proteins expressed in a strain deleted for alg 3 (Δ*alg3* background) were studied for their resistance to Endo-β-N-acetylglucosaminidase H (Endo H) (Aebi, 1996). Previous observations have indicated that only those oligosaccharides larger than $Man_5GlcNAc_2$ are susceptible to cleavage by Endo H (Hubbard, 1980). In the *alg3-1* phenotype, some glycoforms were sensitive to Endo H cleavage, confirming its leakiness, whereas in the Δ*alg3* mutant all glycoforms appeared to be resistant and of the Man_5-type (Aebi, 1996), suggesting a tight phenotype and transfer of $Man_5GlcNAc_2$ oligosaccharide structures onto the nascent polypeptide chain. No obvious phenotype was connected with the inactivation of the *ALG3* gene (Aebi, 1996). Secreted exogluconase produced in a *Saccharomyces cerevisiae alg3* mutant was found to contain between 35-44% underglycosylated and unglycosylated forms and only about 50% of the transferred oligosaccharides remained resistant to Endo H treatment (Cueva, 1996). Exoglucanase (Exg), an enzyme that contains two potential N-glycosylation sites at Asn_{165} and Asn_{325}, was analyzed in more detail. For Exg molecules that received two oligosaccharides it was shown that the first N-glycosylation site (Asn_{165}) was enriched in truncated

residues, whereas the second (Asn_{325}) was enriched in regular oligosaccharides. 35-44% of secreted exoglucanase was non- or underglycosylated and about 73 - 78 % of all available N-glycosylation sites were occupied with either truncated or regular oligosaccharides (Cueva, 1996).

Transfer of Glucosylated Lipid-Linked Oligosaccharides

[0011] Evidence suggests that, in mammalian cells, only glucosylated lipid-linked oligosaccharides are transferred to nascent proteins (Turco, 1977), while in yeast *alg5, alg6* and *dpg1* mutants, nonglucosylated oligosaccharideds can be transferred (Ballou, 1986; Runge, 1984). In a *Saccharomyces cerevisiae alg8* mutant, underglucosylated $GlcMan_9GlcNAc_2$ is transferred (Runge, 1986). Verostek and co-workers studied an *alg3, sec18, gls1* mutant and proposed that glucosylation of a $Man_5GlcNAc_2$ structure (Structure I, above) is relatively slow in comparison to glucosylation of a lipid-linked Man_9 structure. In addition, the transfer of this $Man_5GlcNAc_2$ structure to protein appears to be about 5-fold more efficient than the glucosylation to $Glc_3Man_5GlcNAc_2$. The decreased rate of $Man_5GlcNAc_2$ glucosylation in combination with the comparatively faster rate of Man_5 structure transfer onto nascent protein is believed to be the cause of the observed accumulation of nonglucosylated Man_5 structures in *alg3* mutant yeast (Verostek-a, 1993; Verostek-b, 1993).

[0012] Studies preceding the above work did not reveal any lipid-linked glucosylated oligosaccharides (Orlean, 1990; Huffaker, 1983) allowing the conclusion that glucosylated oligosaccharides are transferred at a much higher rate than their nonglucosylated counterparts and thus are much harder to isolate. Recent work has allowed the creation and study of yeast strains with un- and hypoglucosylated oligosaccharides and has further confirmed the importance of the addition of glucose to the antenna of lipid-linked oligosaccharides for substrate recognition by the oligosaccharyltransferase complex (Reiss, 1996; Stagljar, 1994; Burda, 1998). The decreased degree of glucosylation of the lipid-linked Man_5-oligosaccharides in an *alg3* mutant negatively impacts the kinetics of the transfer of lipid-linked oligosaccharides onto nascent protein and is believed to be the cause for the strong underglycosylation of secreted proteins in an *alg3* knock-out strain (Aebi, 1996).

[0013] The assembly of the lipid-linked core oligosaccharide $Man_9GlcNAc_2$ occurs, as described above, at the membrane of the endoplasmatic reticulum. The additions of three glucose units to the α–1,3-antenna of the lipid-linked oligosaccharides are the final reactions in the oligosaccharide assembly. First an α-1,3 glucose residue is added followed by another α-1,3 glucose residue and a terminal α-1,2 glucose residue. Mutants accumulating dolichol-linked $Man_9GlcNAc_2$ have been shown to be defective in the *ALG6* locus, and Alg6p has similarities to Alg8p, the α-1,3-glucosyltransferase catalyzing the addition of the second α-1,3-linked glucose (Reiss, 1996). Cells with a defective *ALG8* locus accumulate dolichol-linked $Glc_1Man_9GlcNAc_2$ (Runge, 1986; Stagljar, 1994). The *ALG10* locus encodes the α-1,2 glucosyltransferase responsible for the addition of a single terminal glucose to $Glc_2Man_9GlcNAc_2$-PP-Dol (Burda, 1998).

Sequential Processing of N-glycans by Localized Enzyme Activities

[0014] Sugar transferases and mannosidases line the inner (luminal) surface of the ER and Golgi apparatus and thereby provide a "catalytic" surface that allows for the sequential processing of glycoproteins as they proceed through the ER and Golgi network. In fact the multiple compartments of the cis, medial, and trans Golgi and the trans-Golgi Network (TGN), provide the different localities in which the ordered sequence of glycosylation reactions can take place. As a glycoprotein proceeds from synthesis in the ER to full maturation in the late Golgi or TGN, it is sequentially exposed to different glycosidases, mannosidases and glycosyltransferases such that a specific carbohydrate structure may synthesized. Much work has been dedicated to revealing the exact mechanism by which these enzymes are retained and anchored to their respective organelle. The evolving picture is complex but evidence suggests that, stem region, membrane spanning region and cytoplasmic tail individually or in concert direct enzymes to the membrane of individual organelles and thereby localize the associated catalytic domain to that locus.

[0015] In some cases these specific interactions were found to function across species. For example the membrane spanning domain of α2,6-ST from rats, an enzyme known to localize in the trans-Golgi of the animal, was shown to also localize a reporter gene (invertase) in the yeast Golgi (Schwientek, 1995).

However, the very same membrane spanning domain as part of a full-length α2,6 ST was retained in the ER and not further transported to the Golgi of yeast (Krezdorn, 1994). A full length Gal-Tr from humans was not even synthesized in yeast, despite demonstrably high transcription levels. On the other hand the transmembrane region of human the same GalT fused to an invertase reporter was able to direct localization to the yeast Golgi, albeit it at low production levels. Schwientek and co-workers have shown that fusing 28 amino acids of a yeast mannosyltransferase (Mnt1), a region containing a cytoplamic tail, a transmembrane region and eight amino acids of the stem region, to the catalytic domain of human GalT are sufficient for Golgi localization of an active GalT. Other galactosyltransferases appear to rely on interactions with enzymes resident in particular organelles since after removal of their transmembrane region they are still able to localize properly. To date there exists no reliable way of predicting whether a particular heterologously expressed glycosyltransferase or mannosidase in a lower eukaryote will be (1), sufficiently translated (2), catalytically active or (3) located to the proper organelle within the secretory pathway. Since all three of these are necessary to effect glycosylation patterns in lower eukaryotes, a systematic scheme to achieve the desired catalytic function and proper retention of enzymes in the absence of predictive tools, which are currently not available, has been designed.

Production of Therapeutic Glycoproteins

[0016] A significant number of proteins isolated from humans or animals are post-translationally modified, with glycosylation being one of the most significant modifications. An estimated 70% of all therapeutic proteins are glycosylated and thus currently rely on a production system (i.e., host cell) that is able to glycosylate in a manner similar to humans. To date, most glycoproteins are made in a mammalian host system. Several studies have shown that glycosylation plays an important role in determining the (1) immunogenicity, (2) pharmacokinetic properties, (3) trafficking, and (4) efficacy of therapeutic proteins. It is thus not surprising that substantial efforts by the pharmaceutical industry have been directed at developing processes to obtain glycoproteins that are as "humanoid" or "human-like" as possible. This may involve the genetic engineering of such

mammalian cells to enhance the degree of sialylation (i.e., terminal addition of sialic acid) of proteins expressed by the cells, which is known to improve pharmacokinetic properties of such proteins. Alternatively one may improve the degree of sialylation by *in vitro* addition of such sugars using known glycosyltransferases and their respective nucleotide sugars (e.g., 2,3 sialyltransferase and CMP-Sialic acid).

[0017] Future research may reveal the biological and therapeutic significance of specific glycoforms, thereby rendering the ability to produce such specific glycoforms desirable. To date, efforts have concentrated on making proteins with fairly well characterized glycosylation patterns, and expressing a cDNA encoding such a protein in one of the following higher eukaryotic protein expression systems:

1. Higher eukaryotes such as Chinese hamster ovary cells (CHO), mouse fibroblast cells and mouse myeloma cells (Werner, 1998);
2. Transgenic animals such as goats, sheep, mice and others (Dente, 1988); (Cole, 1994); (McGarvey, 1995); (Bardor, 1999);
3. Plants (*Arabidopsis thaliana*, tobacco etc.) (Staub, 2000); (McGarvey, 1995); (Bardor, 1999);
4. Insect cells (*Spodoptera frugiperda* Sf9, Sf21, *Trichoplusia ni*, etc., in combination with recombinant baculoviruses such as *Autographa californica* multiple nuclear polyhedrosis virus which infects lepidopteran cells (Altmann, 1999).

[0018] While most higher eukaryotes carry out glycosylation reactions that are similar to those found in humans, recombinant human proteins expressed in the above mentioned host systems invariably differ from their "natural" human counterpart (Raju, 2000). Extensive development work has thus been directed at finding ways to improving the "human character" of proteins made in these expression systems. This includes the optimization of fermentation conditions and the genetic modification of protein expression hosts by introducing genes encoding enzymes involved in the formation of human like glycoforms (Werner, 1998); (Weikert, 1999); (Andersen, 1994); (Yang, 2000). Inherent problems associated with all mammalian expression systems have not been solved.

[0019] Fermentation processes based on mammalian cell culture (e.g., CHO, murine, or human cells), for example, tend to be very slow (fermentation times in excess of one week are not uncommon), often yield low product titers, require expensive nutrients and cofactors (e.g., bovine fetal serum), are limited by programmed cell death (apoptosis), and often do not enable expression of particular therapeutically valuable proteins. More importantly, mammalian cells are susceptible to viruses that have the potential to be human pathogens and stringent quality controls are required to assure product safety. This is of particular concern since many such processes require the addition of complex and temperature sensitive media components that are derived from animals (e.g., bovine calf serum), which may carry agents pathogenic to humans such as bovine spongiform encephalopathy (BSE) prions or viruses. Moreover, the production of therapeutic compounds is preferably carried out in a well-controlled sterile environment. An animal farm, no matter how cleanly kept, does not constitute such an environment, thus constituting an additional problem in the use of transgenic animals for manufacturing high volume therapeutic proteins.

[0020] Most, if not all, currently produced therapeutic glycoproteins are therefore expressed in mammalian cells and much effort has been directed at improving (i.e., "humanizing") the glycosylation pattern of these recombinant proteins. Changes in medium composition as well as the co-expression of genes encoding enzymes involved in human glycosylation have been successfully employed (see, for example, Weikert, 1999).

[0021] While recombinant proteins similar to their human counterparts can be made in mammalian expression systems, it is currently not possible to make proteins with a human-like glycosylation pattern in lower eukaryotes (fungi and yeast). Although the core oligosaccharide structure transferred to a protein in the endoplasmic reticulum is basically identical in mammals and lower eukaryotes, substantial differences have been found in the subsequent processing reactions which occur in in the Golgi apparatus of fungi and mammals. In fact, even amongst different lower eukaryotes there exist a great variety of glycosylation structures. This has prevented the use of lower eukaryotes as hosts for the production of recombinant human glycoproteins despite otherwise notable

advantages over mammalian expression systems, such as: (1) generally higher product titers, (2) shorter fermentation times, (3) having an alternative for proteins that are poorly expressed in mammalian cells, (4) the ability to grow in a chemically defined protein free medium and thus not requiring complex animal derived media components, (5) and the absence of viral, especially retroviral infections of such hosts.

[0022] Various methylotrophic yeasts such as *Pichia pastoris, Pichia methanolica,* and *Hansenula polymorpha*, have played particularly important roles as eukaryotic expression systems because they are able to grow to high cell densities and secrete large quantities of recombinant protein. However, as noted above, lower eukaryotes such as yeast do not glycosylate proteins like higher mammals. See for example, Martinet *et al.* (1998) Biotechnol Let. Vol. 20. No.12, which discloses the expression of a heterologous mannosidase in the endoplasmic reticulum (ER).

[0023] Chiba et al. (1998) have shown that *S.cerevisiae* can be engineered to provide structures ranging from $Man_8GlcNAc_2$ to $Man_5GlcNAc_2$ structures, by eliminating 1,6 mannosyltransferase (*OCH1*), 1,3 mannosyltransferase (*MNN1*) and a regulator of mannosylphosphatetransferase (*MNN4*) and by targeting the catalytic domain of α-1,2-mannosidase I from *Aspergillus saitoi* into the ER of *S.cerevisiae* using an ER retrieval sequence (Chiba, 1998). However, this attempt resulted in little or no production of the desired $Man_5GlcNAc_2$, e.g., one that was made *in vivo* and which could function as a substrate for GnT1 (the next step in making human-like glycan structures). Chiba et al. (1998) showed that *P. pastoris* is not inherently able to produce useful quantities (greater than 5%) of GlcNAcTransferase I accepting carbohydrate.

[0024] Maras and co-workers assert that in *T. reesei* "sufficient concentrations of acceptor substrate (i.e. $Man_5GlcNAc_2$) are present", however when trying to convert this acceptor substrate to $GlcNAcMan_5GlcNAc_2$ *in vitro* less than 2% were converted thereby demonstrating the presence of $Man_5GlcNAc_2$ structures that are not suitable precursors for complex N-glycan formation (Maras, 1997; Maras, 1999). To date no enabling disclosure exists, that allows for the production of commercially relevant quantities of $GlcNAcMan_5GlcNAc_2$ in lower eukaryotes.

[0025] It is therefore an object of the present invention to provide a system and methods for humanizing glycosylation of recombinant glycoproteins expressed in non-human host cells.

SUMMARY OF THE INVENTION

[0026] The present invention relates to host cells such as fungal strains having modified lipid-linked oligosaccharides which may be modified further by heterologous expression of a set of glycosyltransferases, sugar transporters and mannosidases to become host-strains for the production of mammalian, e.g., human therapeutic glycoproteins. A protein production method has been developed using (1) a lower eukaryotic host such as a unicellular or filamentous fungus, or (2) any non-human eukaryotic organism that has a different glycosylation pattern from humans, to modify the glycosylation composition and structures of the proteins made in a host organism ("host cell") so that they resemble more closely carbohydrate structures found in human proteins. The process allows one to obtain an engineered host cell which can be used to express and target any desirable gene(s) involved in glycosylation by methods that are well established in the scientific literature and generally known to the artisan in the field of protein expression. As described herein, host cells with modified lipid-linked oligosaccharides are created or selected. N-glycans made in the engineered host cells have a $GlcNAcMan_3GlcNAc_2$ core structure which may then be modified further by heterologous expression of one or more enzymes, e.g., glycosyl-transferases, sugar transporters and mannosidases, to yield human-like glycoproteins. For the production of therapeutic proteins, this method may be adapted to engineer cell lines in which any desired glycosylation structure may be obtained.

BRIEF DESCRIPTION OF THE DRAWINGS

[0027] **Figure 1** is a schematic of the structure of the dolichyl pyrophosphate-linked oligosaccharide.

[0028] **Figure 2** is a schematic of the generation of $GlcNAc_2Man_3GlcNAc_2$ N-glycans from fungal host cells which are deficient in alg3, alg9 or alg 12 activities.

[0029] **Figure 3** is a schematic of processing reactions required to produce mammalian-type oligosaccharide structures in a fungal host cell with an *alg3*, *och1* genotype.

[0030] **Figure 4** shows *S. cerevisiae Alg3* Sequence Comparisons (Blast)

[0031] **Figure 5** shows *S. cerevisiae Alg 3* and Alg 3p Sequences

[0032] **Figure 6** shows *P. pastoris Alg 3* and Alg 3p Sequences

[0033] **Figure 7** shows *P. pastoris Alg 3* Sequence Comparisons (Blast)

[0034] **Figure 8** shows *K. lactis Alg 3* and Alg 3p Sequences

[0035] **Figure 9** shows *K. lactis Alg 3* Sequence Comparisons (Blast)

[0036] **Figure 10** shows *S. cerevisiae Alg 9* and Alg 9p Sequences

[0037] **Figure 11** shows *P. pastoris Alg 9* and Alg 9p Sequences

[0038] **Figure 12** shows *P. pastoris Alg 9* Sequence Comparisons (Blast)

[0039] **Figure 13** shows *S. cerevisiae Alg 12* and Alg 12p Sequences

[0040] **Figure 14** shows *P. pastoris Alg 12* and Alg 12p Sequences

[0041] **Figure 15** shows *P. pastoris Alg 12* Sequence Comparisons (Blast)

[0042] **Figure 16** is a MALDI-TOF-MS analysis of N-glycans isolated from a kringle 3 glycoprotein produced in a *P.pastoris* showing that the predominant N-glycan is $GlcNAcMan_5GlcNAc_2$.

[0043] **Figure 17** is a MALDI-TOF-MS analysis of N-glycans isolated from a kringle 3 glycoprotein produced in a *P.pastoris* (Fig. 16) treated with β-N-hexosaminidase (peak corresponding to $Man_5GlcNAc_2$) to confirm that the predominant N-glycan of Fig. 16 is $GlcNAcMan_5GlcNAc_2$.

[0044] **Figure 18** is a MALDI-TOF-MS analysis of N-glycans isolated from a kringle 3 glycoprotein produced in a *P.pastoris alg3* deletion mutant showing that the predominant N-glycans are $GlcNAcMan_3GlcNAc_2$ and $GlcNAcMan_4GlcNAc_2$.

[0045] **Figure 19** is a MALDI-TOF-MS analysis of N-glycans isolated from a kringle 3 glycoprotein produced in a *P.pastoris alg3* deletion mutant treated with α1,2 mannosidase, showing that the $GlcNAcMan_4GlcNAc_2$ of Fig. 18 is converted to $GlcNAcMan_3GlcNAc_2$.

[0046] **Figure 20** is a MALDI-TOF-MS analysis of N-glycans of Fig. 19 treated with β-N-hexosaminidase (peak corresponding to $Man_3GlcNAc_2$) to confirm that the N-glycan of Fig. 19 is $GlcNAcMan_3GlcNAc_2$.

[0047] **Figure 21** is a MALDI-TOF-MS analysis of N-glycans isolated from a kringle 3 glycoprotein produced in a *P.pastoris alg3* deletion mutant treated with α1,2 mannosidase and GnTII, showing that the $GlcNAcMan_3GlcNAc_2$ of Fig. 19 is converted to $GlcNAc_2Man_3GlcNAc_2$.

[0048] **Figure 22** is a MALDI-TOF-MS analysis of N-glycans of Fig. 21 treated with β-N-hexosaminidase (peak corresponding to $Man_3GlcNAc_2$) to confirm that the N-glycan of Fig. 21 is $GlcNAc_2Man_3GlcNAc_2$.

[0049] **Figure 23** is a MALDI-TOF-MS analysis of N-glycans isolated from a kringle 3 glycoprotein produced in a *P.pastoris alg3* deletion mutant treated with α1,2 mannosidase and GnTII in the presence of UDP-galactose and β1,4-galactosyltransferase, showing that the $GlcNAc_2Man_3GlcNAc_2$ of Fig. 21 is converted to $Gal_2GlcNAc_2Man_3GlcNAc_2$.

[0050] **Figure 24** is a MALDI-TOF-MS analysis of N-glycans isolated from a kringle 3 glycoprotein produced in a *P.pastoris alg3* deletion mutant treated with α1,2 mannosidase and GnTII in the presence of UDP-galactose and β1,4-galactosyltransferase, and further treated with CMP-N-acetylneuraminic acid and sialyltransferase, showing that the $Gal_2GlcNAc_2Man_3GlcNAc_2$ is converted to $NANA_2Gal_2GlcNAc_2Man_3GlcNAc_2$.

[0051] **Figure 25** shows *S. cerevisiae Alg6* and Alg 6p Sequences

[0052] **Figure 26** shows *P. pastoris Alg6* and Alg 6p Sequences

[0053] **Figure 27** shows *P. pastoris Alg 6* Sequence Comparisons (Blast)

[0054] **Figure 28**shows *K.lactis Alg6* and Alg 6p Sequences

[0055] **Figure 29** shows *K.lactis Alg 6* Sequence Comparisons (Blast)

[0056] **Figure 30** Model of an IgG immunoglobulin. Heavy chain and light chain can be, based on similar secondary and tertiary structure, subdivided into domains. The two heavy chains (domains V_H, C_H1, C_H2 and C_H3) are linked through three disulfide bridges. The light chains (domains V_L and C_L) are linked by another disulfide bridge to the C_H1 portion of the heavy chain and, together with the C_H1 and V_H fragments, make up the Fab region. Antigens bind to the terminal

portion of the Fab region. Effector-functions, such as Fc-gamma-Receptor binding have been localized to the C_H2 domain, just downstream of the hinge region and are influenced by N-glycosylation of asparagine 297 in the heavy chain.

[0057] **Figure 31** Schematic overview of a modular IgG1 expression vector.

[0058] **Figure 32** shows *M. musculis GnT III* Nucleic Acid And Amino Acid Sequences

[0059] **Figure 33** shows *H. sapiens GnT IV* Nucleic Acid And Amino Acid Sequences

[0060] **Figure 34** shows *M. musculis GnT V* Nucleic Acid And Amino Acid Sequences

DETAILED DESCRIPTION OF THE INVENTION

[0061] Unless otherwise defined herein, scientific and technical terms used in connection with the present invention shall have the meanings that are commonly understood by those of ordinary skill in the art. Further, unless otherwise required by context, singular terms shall include pluralities and plural terms shall include the singular. The methods and techniques of the present invention are generally performed according to conventional methods well known in the art. Generally, nomenclatures used in connection with, and techniques of biochemistry, enzymology, molecular and cellular biology, microbiology, genetics and protein and nucleic acid chemistry and hybridization described herein are those well known and commonly used in the art. The methods and techniques of the present invention are generally performed according to conventional methods well known in the art and as described in various general and more specific references that are cited and discussed throughout the present specification unless otherwise indicated. See, e.g., Sambrook et al. Molecular Cloning: A Laboratory Manual, 2d ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989); Ausubel et al., Current Protocols in Molecular Biology, Greene Publishing Associates (1992, and Supplements to 2002); Harlow and Lane Antibodies: A Laboratory Manual Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1990); Introduction to Glycobiology, Maureen E. Taylor, Kurt Drickamer, Oxford Univ. Press (2003); Worthington Enzyme Manual, Worthington Biochemical Corp. Freehold, NJ;

Handbook of Biochemistry: Section A Proteins Vol I 1976 CRC Press; Handbook of Biochemistry: Section A Proteins Vol II 1976 CRC Press; Essentials of Glycobiology, Cold Spring Harbor Laboratory Press (1999). The nomenclatures used in connection with, and the laboratory procedures and techniques of, biochemistry and molecular biology described herein are those well known and commonly used in the art.

[0062] All publications, patents and other references mentioned herein are incorporated by reference.

[0063] The following terms, unless otherwise indicated, shall be understood to have the following meanings:

[0064] As used herein, the term "N-glycan" refers to an N-linked oligosaccharide, e.g., one that is attached by an asparagine-N-acetylglucosamine linkage to an asparagine residue of a polypeptide. N-glycans have a common pentasaccharide core of $Man_3GlcNAc_2$ ("Man" refers to mannose; "Glc" refers to glucose; and "NAc" refers to N-acetyl; GlcNAc refers to N-acetylglucosamine). N-glycans differ with respect to the number of branches (antennae) comprising peripheral sugars (e.g., fucose and sialic acid) that are added to the $Man_3GlcNAc_2$ ("Man3") core structure. N-glycans are classified according to their branched constituents (e.g., high mannose, complex or hybrid). A "high mannose" type N-glycan has five or more mannose residues. A "complex" type N-glycan typically has at least one GlcNAc attached to the 1,3 mannose arm and at least one GlcNAc attached to the 1,6 mannose arm of a "trimannose" core. The "trimannose core" is the pentasaccharide core having a Man3 structure. Complex N-glycans may also have galactose ("Gal") residues that are optionally modified with sialic acid or derivatives ("NeuAc", where "Neu" refers to neuraminic acid and "Ac" refers to acetyl). Complex N-glycans may also have intrachain substitutions comprising "bisecting" GlcNAc and core fucose ("Fuc"). A "hybrid" N-glycan has at least one GlcNAc on the terminal of the 1,3 mannose arm of the trimannose core and zero or more mannoses on the 1,6 mannose arm of the trimannose core.

[0065] Abbreviations used herein are of common usage in the art, see, e.g., abbreviations of sugars, above. Other common abbreviations include "PNGase", which refers to peptide N-glycosidase F (EC 3.2.2.18); "GlcNAc Tr (I – III)",

which refers to one of three N-acetylglucosaminyltransferase enzymes; "NANA" refers to N-acetylneuraminic acid.

[0066] As used herein, the term "secretion pathway" refers to the assembly line of various glycosylation enzymes to which a lipid-linked oligosaccharide precursor and an N-glycan substrate are sequentially exposed, following the molecular flow of a nascent polypeptide chain from the cytoplasm to the endoplasmic reticulum (ER) and the compartments of the Golgi apparatus. Enzymes are said to be localized along this pathway. An enzyme X that acts on a lipid-linked glycan or an N-glycan before enzyme Y is said to be or to act "upstream" to enzyme Y; similarly, enzyme Y is or acts "downstream" from enzyme X.

[0067] As used herein, the term "alg X activity" refers to the enzymatic activity encoded by the "alg X" gene, and to an enzyme having that enzymatic activity encoded by a homologous gene or gene product (see below) or by an unrelated gene or gene product.

[0068] As used herein, the term "antibody" refers to a full antibody (consisting of two heavy chains and two light chains) or a fragment thereof. Such fragments include, but are not limited to, those produced by digestion with various proteases, those produced by chemical cleavage and/or chemical dissociation, and those produced recombinantly, so long as the fragment remains capable of specific binding to an antigen. Among these fragments are Fab, Fab', F(ab')2, and single chain Fv (scFv) fragments. Within the scope of the term "antibody" are also antibodies that have been modified in sequence, but remain capable of specific binding to an antigen. Example of modified antibodies are interspecies chimeric and humanized antibodies; antibody fusions; and heteromeric antibody complexes, such as diabodies (bispecific antibodies), single-chain diabodies, and intrabodies (see, e.g., Marasco (ed.), Intracellular Antibodies: Research and Disease Applications, Springer-Verlag New York, Inc. (1998) (ISBN: 3540641513), the disclosure of which is incorporated herein by reference in its entirety).

[0069] As used herein, the term "mutation" refers to any change in the nucleic acid or amino acid sequence of a gene product, e.g., of a glycosylation-related enzyme.

[0070] The term "polynucleotide" or "nucleic acid molecule" refers to a polymeric form of nucleotides of at least 10 bases in length. The term includes DNA molecules (e.g., cDNA or genomic or synthetic DNA) and RNA molecules (e.g., mRNA or synthetic RNA), as well as analogs of DNA or RNA containing non-natural nucleotide analogs, non-native internucleoside bonds, or both. The nucleic acid can be in any topological conformation. For instance, the nucleic acid can be single-stranded, double-stranded, triple-stranded, quadruplexed, partially double-stranded, branched, hairpinned, circular, or in a padlocked conformation. The term includes single and double stranded forms of DNA.

[0071] Unless otherwise indicated, a "nucleic acid comprising SEQ ID NO:X" refers to a nucleic acid, at least a portion of which has either (i) the sequence of SEQ ID NO:X, or (ii) a sequence complementary to SEQ ID NO:X. The choice between the two is dictated by the context. For instance, if the nucleic acid is used as a probe, the choice between the two is dictated by the requirement that the probe be complementary to the desired target.

[0072] An "isolated" or "substantially pure" nucleic acid or polynucleotide (e.g., an RNA, DNA or a mixed polymer) is one which is substantially separated from other cellular components that naturally accompany the native polynucleotide in its natural host cell, e.g., ribosomes, polymerases, and genomic sequences with which it is naturally associated. The term embraces a nucleic acid or polynucleotide that (1) has been removed from its naturally occurring environment, (2) is not associated with all or a portion of a polynucleotide in which the "isolated polynucleotide" is found in nature, (3) is operatively linked to a polynucleotide which it is not linked to in nature, or (4) does not occur in nature. The term "isolated" or "substantially pure" also can be used in reference to recombinant or cloned DNA isolates, chemically synthesized polynucleotide analogs, or polynucleotide analogs that are biologically synthesized by heterologous systems.

[0073] However, "isolated" does not necessarily require that the nucleic acid or polynucleotide so described has itself been physically removed from its native environment. For instance, an endogenous nucleic acid sequence in the genome of an organism is deemed "isolated" herein if a heterologous sequence (i.e., a sequence that is not naturally adjacent to this endogenous nucleic acid sequence) is

placed adjacent to the endogenous nucleic acid sequence, such that the expression of this endogenous nucleic acid sequence is altered. By way of example, a non-native promoter sequence can be substituted (e.g., by homologous recombination) for the native promoter of a gene in the genome of a human cell, such that this gene has an altered expression pattern. This gene would now become "isolated" because it is separated from at least some of the sequences that naturally flank it.

[0074] A nucleic acid is also considered "isolated" if it contains any modifications that do not naturally occur to the corresponding nucleic acid in a genome. For instance, an endogenous coding sequence is considered "isolated" if it contains an insertion, deletion or a point mutation introduced artificially, e.g., by human intervention. An "isolated nucleic acid" also includes a nucleic acid integrated into a host cell chromosome at a heterologous site, a nucleic acid construct present as an episome. Moreover, an "isolated nucleic acid" can be substantially free of other cellular material, or substantially free of culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized.

[0075] As used herein, the phrase "degenerate variant" of a reference nucleic acid sequence encompasses nucleic acid sequences that can be translated, according to the standard genetic code, to provide an amino acid sequence identical to that translated from the reference nucleic acid sequence.

[0076] The term "percent sequence identity" or "identical" in the context of nucleic acid sequences refers to the residues in the two sequences which are the same when aligned for maximum correspondence. The length of sequence identity comparison may be over a stretch of at least about nine nucleotides, usually at least about 20 nucleotides, more usually at least about 24 nucleotides, typically at least about 28 nucleotides, more typically at least about 32 nucleotides, and preferably at least about 36 or more nucleotides. There are a number of different algorithms known in the art which can be used to measure nucleotide sequence identity. For instance, polynucleotide sequences can be compared using FASTA, Gap or Bestfit, which are programs in Wisconsin Package Version 10.0, Genetics Computer Group (GCG), Madison, Wisconsin. FASTA provides alignments and percent sequence identity of the regions of the best overlap between the query and search

sequences (Pearson, 1990, (herein incorporated by reference). For instance, percent sequence identity between nucleic acid sequences can be determined using FASTA with its default parameters (a word size of 6 and the NOPAM factor for the scoring matrix) or using Gap with its default parameters as provided in GCG Version 6.1, herein incorporated by reference.

[0077] The term "substantial homology" or "substantial similarity," when referring to a nucleic acid or fragment thereof, indicates that, when optimally aligned with appropriate nucleotide insertions or deletions with another nucleic acid (or its complementary strand), there is nucleotide sequence identity in at least about 50%, more preferably 60% of the nucleotide bases, usually at least about 70%, more usually at least about 80%, preferably at least about 90%, and more preferably at least about 95%, 96%, 97%, 98% or 99% of the nucleotide bases, as measured by any well-known algorithm of sequence identity, such as FASTA, BLAST or Gap, as discussed above.

[0078] Alternatively, substantial homology or similarity exists when a nucleic acid or fragment thereof hybridizes to another nucleic acid, to a strand of another nucleic acid, or to the complementary strand thereof, under stringent hybridization conditions. "Stringent hybridization conditions" and "stringent wash conditions" in the context of nucleic acid hybridization experiments depend upon a number of different physical parameters. Nucleic acid hybridization will be affected by such conditions as salt concentration, temperature, solvents, the base composition of the hybridizing species, length of the complementary regions, and the number of nucleotide base mismatches between the hybridizing nucleic acids, as will be readily appreciated by those skilled in the art. One having ordinary skill in the art knows how to vary these parameters to achieve a particular stringency of hybridization.

[0079] In general, "stringent hybridization" is performed at about 25°C below the thermal melting point (T_m) for the specific DNA hybrid under a particular set of conditions. "Stringent washing" is performed at temperatures about 5°C lower than the T_m for the specific DNA hybrid under a particular set of conditions. The T_m is the temperature at which 50% of the target sequence hybridizes to a perfectly matched probe. See Sambrook et al., *supra*, page 9.51, hereby incorporated by

reference. For purposes herein, "high stringency conditions" are defined for solution phase hybridization as aqueous hybridization (i.e., free of formamide) in 6X SSC (where 20X SSC contains 3.0 M NaCl and 0.3 M sodium citrate), 1% SDS at 65oC for 8-12 hours, followed by two washes in 0.2X SSC, 0.1% SDS at 65oC for 20 minutes. It will be appreciated by the skilled worker that hybridization at 65°C will occur at different rates depending on a number of factors including the length and percent identity of the sequences which are hybridizing.

[0080] The nucleic acids (also referred to as polynucleotides) of this invention may include both sense and antisense strands of RNA, cDNA, genomic DNA, and synthetic forms and mixed polymers of the above. They may be modified chemically or biochemically or may contain non-natural or derivatized nucleotide bases, as will be readily appreciated by those of skill in the art. Such modifications include, for example, labels, methylation, substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as uncharged linkages (e.g., methyl phosphonates, phosphotriesters, phosphoramidates, carbamates, etc.), charged linkages (e.g., phosphorothioates, phosphorodithioates, etc.), pendent moieties (e.g., polypeptides), intercalators (e.g., acridine, psoralen, etc.), chelators, alkylators, and modified linkages (e.g., alpha anomeric nucleic acids, etc.) Also included are synthetic molecules that mimic polynucleotides in their ability to bind to a designated sequence via hydrogen bonding and other chemical interactions. Such molecules are known in the art and include, for example, those in which peptide linkages substitute for phosphate linkages in the backbone of the molecule.

[0081] The term "mutated" when applied to nucleic acid sequences means that nucleotides in a nucleic acid sequence may be inserted, deleted or changed compared to a reference nucleic acid sequence. A single alteration may be made at a locus (a point mutation) or multiple nucleotides may be inserted, deleted or changed at a single locus. In addition, one or more alterations may be made at any number of loci within a nucleic acid sequence. A nucleic acid sequence may be mutated by any method known in the art including but not limited to mutagenesis techniques such as "error-prone PCR" (a process for performing PCR under conditions where the copying fidelity of the DNA polymerase is low, such that a

high rate of point mutations is obtained along the entire length of the PCR product. See, e.g., Leung, D. W., et al., *Technique*, 1, pp. 11-15 (1989) and Caldwell, R. C. & Joyce G. F., *PCR Methods Applic.*, 2, pp. 28-33 (1992)); and "oligonucleotide-directed mutagenesis" (a process which enables the generation of site-specific mutations in any cloned DNA segment of interest. See, e.g., Reidhaar-Olson, J. F. & Sauer, R. T., et al., *Science*, 241, pp. 53-57 (1988)).

[0082] The term "vector" as used herein is intended to refer to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments may be ligated. Other vectors include cosmids, bacterial artificial chromosomes (BAC) and yeast artificial chromosomes (YAC). Another type of vector is a viral vector, wherein additional DNA segments may be ligated into the viral genome (discussed in more detail below). Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., vectors having an origin of replication which functions in the host cell). Other vectors can be integrated into the genome of a host cell upon introduction into the host cell, and are thereby replicated along with the host genome. Moreover, certain preferred vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "recombinant expression vectors" (or simply, "expression vectors").

[0083] "Operatively linked" expression control sequences refers to a linkage in which the expression control sequence is contiguous with the gene of interest to control the gene of interest, as well as expression control sequences that act in *trans* or at a distance to control the gene of interest.

[0084] The term "expression control sequence" as used herein refers to polynucleotide sequences which are necessary to affect the expression of coding sequences to which they are operatively linked. Expression control sequences are sequences which control the transcription, post-transcriptional events and translation of nucleic acid sequences. Expression control sequences include appropriate transcription initiation, termination, promoter and enhancer sequences; efficient RNA processing signals such as splicing and polyadenylation signals; sequences that stabilize cytoplasmic mRNA; sequences that enhance translation

efficiency (e.g., ribosome binding sites); sequences that enhance protein stability; and when desired, sequences that enhance protein secretion. The nature of such control sequences differs depending upon the host organism; in prokaryotes, such control sequences generally include promoter, ribosomal binding site, and transcription termination sequence. The term "control sequences" is intended to include, at a minimum, all components whose presence is essential for expression, and can also include additional components whose presence is advantageous, for example, leader sequences and fusion partner sequences.

[0085] The term "recombinant host cell" (or simply "host cell"), as used herein, is intended to refer to a cell into which a recombinant vector has been introduced. It should be understood that such terms are intended to refer not only to the particular subject cell but to the progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term "host cell" as used herein. A recombinant host cell may be an isolated cell or cell line grown in culture or may be a cell which resides in a living tissue or organism.

[0086] The term "peptide" as used herein refers to a short polypeptide, e.g., one that is typically less than about 50 amino acids long and more typically less than about 30 amino acids long. The term as used herein encompasses analogs and mimetics that mimic structural and thus biological function.

[0087] The term "polypeptide" encompasses both naturally-occurring and non-naturally-occurring proteins, and fragments, mutants, derivatives and analogs thereof. A polypeptide may be monomeric or polymeric. Further, a polypeptide may comprise a number of different domains each of which has one or more distinct activities.

[0088] The term "isolated protein" or "isolated polypeptide" is a protein or polypeptide that by virtue of its origin or source of derivation (1) is not associated with naturally associated components that accompany it in its native state, (2) when it exists in a purity not found in nature, where purity can be adjudged with respect to the presence of other cellular material (e.g., is free of other proteins from the same species) (3) is expressed by a cell from a different species, or (4) does not

occur in nature (e.g., it is a fragment of a polypeptide found in nature or it includes amino acid analogs or derivatives not found in nature or linkages other than standard peptide bonds). Thus, a polypeptide that is chemically synthesized or synthesized in a cellular system different from the cell from which it naturally originates will be "isolated" from its naturally associated components. A polypeptide or protein may also be rendered substantially free of naturally associated components by isolation, using protein purification techniques well known in the art. As thus defined, "isolated" does not necessarily require that the protein, polypeptide, peptide or oligopeptide so described has been physically removed from its native environment.

[0089] The term "polypeptide fragment" as used herein refers to a polypeptide that has an amino-terminal and/or carboxy-terminal deletion compared to a full-length polypeptide. In a preferred embodiment, the polypeptide fragment is a contiguous sequence in which the amino acid sequence of the fragment is identical to the corresponding positions in the naturally-occurring sequence. Fragments typically are at least 5, 6, 7, 8, 9 or 10 amino acids long, preferably at least 12, 14, 16 or 18 amino acids long, more preferably at least 20 amino acids long, more preferably at least 25, 30, 35, 40 or 45, amino acids, even more preferably at least 50 or 60 amino acids long, and even more preferably at least 70 amino acids long.

[0090] A "modified derivative" refers to polypeptides or fragments thereof that are substantially homologous in primary structural sequence but which include, e.g., *in vivo* or *in vitro* chemical and biochemical modifications or which incorporate amino acids that are not found in the native polypeptide. Such modifications include, for example, acetylation, carboxylation, phosphorylation, glycosylation, ubiquitination, labeling, e.g., with radionuclides, and various enzymatic modifications, as will be readily appreciated by those well skilled in the art. A variety of methods for labeling polypeptides and of substituents or labels useful for such purposes are well known in the art, and include radioactive isotopes such as ^{125}I, ^{32}P, ^{35}S, and ^{3}H, ligands which bind to labeled antiligands (e.g., antibodies), fluorophores, chemiluminescent agents, enzymes, and antiligands which can serve as specific binding pair members for a labeled ligand. The choice of label depends on the sensitivity required, ease of conjugation with the primer,

stability requirements, and available instrumentation. Methods for labeling polypeptides are well known in the art. See Ausubel et al., 1992, hereby incorporated by reference.

[0091] The term "fusion protein" refers to a polypeptide comprising a polypeptide or fragment coupled to heterologous amino acid sequences. Fusion proteins are useful because they can be constructed to contain two or more desired functional elements from two or more different proteins. A fusion protein comprises at least 10 contiguous amino acids from a polypeptide of interest, more preferably at least 20 or 30 amino acids, even more preferably at least 40, 50 or 60 amino acids, yet more preferably at least 75, 100 or 125 amino acids. Fusion proteins can be produced recombinantly by constructing a nucleic acid sequence which encodes the polypeptide or a fragment thereof in frame with a nucleic acid sequence encoding a different protein or peptide and then expressing the fusion protein. Alternatively, a fusion protein can be produced chemically by crosslinking the polypeptide or a fragment thereof to another protein.

[0092] The term "non-peptide analog" refers to a compound with properties that are analogous to those of a reference polypeptide. A non-peptide compound may also be termed a "peptide mimetic" or a "peptidomimetic". See, e.g., Jones, (1992) Amino Acid and Peptide Synthesis, Oxford University Press; Jung, (1997) Combinatorial Peptide and Nonpeptide Libraries: A Handbook John Wiley; Bodanszky et al., (1993) Peptide Chemistry--A Practical Textbook, Springer Verlag; "Synthetic Peptides: A Users Guide", G. A. Grant, Ed, W. H. Freeman and Co., 1992; Evans et al. *J. Med. Chem.* 30:1229 (1987); Fauchere, *J. Adv. Drug Res.* 15:29 (1986); Veber and Freidinger *TINS* p.392 (1985); and references sited in each of the above, which are incorporated herein by reference. Such compounds are often developed with the aid of computerized molecular modeling. Peptide mimetics that are structurally similar to useful peptides of the invention may be used to produce an equivalent effect and are therefore envisioned to be part of the invention.

[0093] A "polypeptide mutant" or "mutein" refers to a polypeptide whose sequence contains an insertion, duplication, deletion, rearrangement or substitution of one or more amino acids compared to the amino acid sequence of a native or

wild type protein. A mutein may have one or more amino acid point substitutions, in which a single amino acid at a position has been changed to another amino acid, one or more insertions and/or deletions, in which one or more amino acids are inserted or deleted, respectively, in the sequence of the naturally-occurring protein, and/or truncations of the amino acid sequence at either or both the amino or carboxy termini. A mutein may have the same but preferably has a different biological activity compared to the naturally-occurring protein. For instance, a mutein may have an increased or decreased neuron or NgR binding activity. In a preferred embodiment of the present invention, a MAG derivative that is a mutein (e.g., in MAG Ig-like domain 5) has decreased neuronal growth inhibitory activity compared to endogenous or soluble wild-type MAG.

[0094] A mutein has at least 70% overall sequence homology to its wild-type counterpart. Even more preferred are muteins having 80%, 85% or 90% overall sequence homology to the wild-type protein. In an even more preferred embodiment, a mutein exhibits 95% sequence identity, even more preferably 97%, even more preferably 98% and even more preferably 99% overall sequence identity. Sequence homology may be measured by any common sequence analysis algorithm, such as Gap or Bestfit.

[0095] Preferred amino acid substitutions are those which: (1) reduce susceptibility to proteolysis, (2) reduce susceptibility to oxidation, (3) alter binding affinity for forming protein complexes, (4) alter binding affinity or enzymatic activity, and (5) confer or modify other physicochemical or functional properties of such analogs.

[0096] As used herein, the twenty conventional amino acids and their abbreviations follow conventional usage. *See Immunology - A Synthesis* (2nd Edition, E.S. Golub and D.R. Gren, Eds., Sinauer Associates, Sunderland, Mass. (1991)), which is incorporated herein by reference. Stereoisomers (e.g., D-amino acids) of the twenty conventional amino acids, unnatural amino acids such as α-, α-disubstituted amino acids, N-alkyl amino acids, and other unconventional amino acids may also be suitable components for polypeptides of the present invention. Examples of unconventional amino acids include: 4-hydroxyproline, γ-carboxyglutamate, ϵ-N,N,N-trimethyllysine, ϵ-N-acetyllysine, O-phosphoserine,

N-acetylserine, N-formylmethionine, 3-methylhistidine, 5-hydroxylysine, s-N-methylarginine, and other similar amino acids and imino acids (e.g., 4-hydroxyproline). In the polypeptide notation used herein, the left-hand direction is the amino terminal direction and the right hand direction is the carboxy-terminal direction, in accordance with standard usage and convention.

[0097] A protein has "homology" or is "homologous" to a second protein if the nucleic acid sequence that encodes the protein has a similar sequence to the nucleic acid sequence that encodes the second protein. Alternatively, a protein has homology to a second protein if the two proteins have "similar" amino acid sequences. (Thus, the term "homologous proteins" is defined to mean that the two proteins have similar amino acid sequences). In a preferred embodiment, a homologous protein is one that exhibits 60% sequence homology to the wild type protein, more preferred is 70% sequence homology. Even more preferred are homologous proteins that exhibit 80%, 85% or 90% sequence homology to the wild type protein. In a yet more preferred embodiment, a homologous protein exhibits 95%, 97%, 98% or 99% sequence identity. As used herein, homology between two regions of amino acid sequence (especially with respect to predicted structural similarities) is interpreted as implying similarity in function.

[0098] When "homologous" is used in reference to proteins or peptides, it is recognized that residue positions that are not identical often differ by conservative amino acid substitutions. A "conservative amino acid substitution" is one in which an amino acid residue is substituted by another amino acid residue having a side chain (R group) with similar chemical properties (e.g., charge or hydrophobicity). In general, a conservative amino acid substitution will not substantially change the functional properties of a protein. In cases where two or more amino acid sequences differ from each other by conservative substitutions, the percent sequence identity or degree of homology may be adjusted upwards to correct for the conservative nature of the substitution. Means for making this adjustment are well known to those of skill in the art (see, e.g., Pearson et al., 1994, herein incorporated by reference).

[0099] The following six groups each contain amino acids that are conservative substitutions for one another: 1) Serine (S), Threonine (T); 2) Aspartic Acid (D),

Glutamic Acid (E); 3) Asparagine (N), Glutamine (Q); 4) Arginine (R), Lysine (K); 5) Isoleucine (I), Leucine (L), Methionine (M), Alanine (A), Valine (V), and 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

[0100] Sequence homology for polypeptides, which is also referred to as percent sequence identity, is typically measured using sequence analysis software. See, e.g., the Sequence Analysis Software Package of the Genetics Computer Group (GCG), University of Wisconsin Biotechnology Center, 910 University Avenue, Madison, Wisconsin 53705. Protein analysis software matches similar sequences using measure of homology assigned to various substitutions, deletions and other modifications, including conservative amino acid substitutions. For instance, GCG contains programs such as "Gap" and "Bestfit" which can be used with default parameters to determine sequence homology or sequence identity between closely related polypeptides, such as homologous polypeptides from different species of organisms or between a wild type protein and a mutein thereof. See, e.g., GCG Version 6.1.

[0101] A preferred algorithm when comparing a inhibitory molecule sequence to a database containing a large number of sequences from different organisms is the computer program BLAST (Altschul, S.F. et al. (1990) *J. Mol. Biol.* 215:403-410; Gish and States (1993) *Nature Genet.* 3:266-272; Madden, T.L. et al. (1996) *Meth. Enzymol.* 266:131-141; Altschul, S.F. et al. (1997) *Nucleic Acids Res.*25:3389-3402; Zhang, J. and Madden, T.L. (1997) *Genome Res.* 7:649-656), especially blastp or tblastn (Altschul et al., 1997). Preferred parameters for BLASTp are:

Expectation value:	10 (default)
Filter:	seg (default)
Cost to open a gap:	11 (default)
Cost to extend a gap:	1 (default
Max. alignments:	100 (default)
Word size:	11 (default)
No. of descriptions:	100 (default)
Penalty Matrix:	BLOWSUM62

[0102] The length of polypeptide sequences compared for homology will generally be at least about 16 amino acid residues, usually at least about 20

residues, more usually at least about 24 residues, typically at least about 28 residues, and preferably more than about 35 residues. When searching a database containing sequences from a large number of different organisms, it is preferable to compare amino acid sequences. Database searching using amino acid sequences can be measured by algorithms other than blastp known in the art. For instance, polypeptide sequences can be compared using FASTA, a program in GCG Version 6.1. FASTA provides alignments and percent sequence identity of the regions of the best overlap between the query and search sequences (Pearson, 1990, herein incorporated by reference). For example, percent sequence identity between amino acid sequences can be determined using FASTA with its default parameters (a word size of 2 and the PAM250 scoring matrix), as provided in GCG Version 6.1, herein incorporated by reference.

[0103] "Specific binding" refers to the ability of two molecules to bind to each other in preference to binding to other molecules in the environment. Typically, "specific binding" discriminates over adventitious binding in a reaction by at least two-fold, more typically by at least 10-fold, often at least 100-fold. Typically, the affinity or avidity of a specific binding reaction is at least about 10-7 M (e.g., at least about 10^{-8} M or 10^{-9} M).

[0104] The term "region" as used herein refers to a physically contiguous portion of the primary structure of a biomolecule. In the case of proteins, a region is defined by a contiguous portion of the amino acid sequence of that protein.

[0105] The term "domain" as used herein refers to a structure of a biomolecule that contributes to a known or suspected function of the biomolecule. Domains may be co-extensive with regions or portions thereof; domains may also include distinct, non-contiguous regions of a biomolecule. Examples of protein domains include, but are not limited to, an Ig domain, an extracellular domain, a transmembrane domain, and a cytoplasmic domain.

[0106] As used herein, the term "molecule" means any compound, including, but not limited to, a small molecule, peptide, protein, sugar, nucleotide, nucleic acid, lipid, etc., and such a compound can be natural or synthetic.

[0107] Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art

to which this invention pertains. Exemplary methods and materials are described below, although methods and materials similar or equivalent to those described herein can also be used in the practice of the present invention and will be apparent to those of skill in the art. All publications and other references mentioned herein are incorporated by reference in their entirety. In case of conflict, the present specification, including definitions, will control. The materials, methods, and examples are illustrative only and not intended to be limiting.

[0108] Throughout this specification and claims, the word "comprise" or variations such as "comprises" or "comprising", will be understood to imply the inclusion of a stated integer or group of integers but not the exclusion of any other integer or group of integers.

Engineering or Selecting Hosts With Modified Lipid-Linked Oligosaccharides For The Generation of Human-like N-Glycans

[0109] The invention provides a method for producing a human-like glycoprotein in a non-human eukaryotic host cell. The method involves making or using a non-human eukaryotic host cell diminished or depleted in an *alg* gene activity (i.e., *alg* activities, including equivalent enzymatic activities in non-fungal host cells) and introducing into the host cell at least one glycosidase activity. In a preferred embodiment, the glycosidase activity is introduced by causing expression of one or more mannosidase activities within the host cell, for example, by activation of a mannosidase activity, or by expression from a nucleic acid molecule of a mannosidase activity, in the host cell.

[0110] In another embodiment, the method involves making or using a host cell diminished or depleted in the activity of one or more enzymes that transfer a sugar residue to the 1,6 arm of lipid-linked oligosaccharide precursors (**Fig. 1**). A host cell of the invention is selected for or is engineered by introducing a mutation in one or more of the genes encoding an enzyme that transfers a sugar residue (e.g., mannosylates) the 1,6 arm of a lipid-linked oligosaccharide precursor. The sugar residue is more preferably mannose, is preferably a glucose, GlcNAc, galactose, sialic acid, fucose or GlcNAc phosphate residue. In a preferred embodiment, the activity of one or more enzymes that mannosylate the 1,6 arm of lipid-linked

oligosaccharide precursors is diminished or depleted. The method may further comprise the step of introducing into the host cell at least one glycosidase activity (see below).

[0111] In yet another embodiment, the invention provides a method for producing a human-like glycoprotein in a non-human host, wherein the glycoprotein comprises an N-glycan having at least two GlcNAcs attached to a trimannose core structure.

[0112] In each above embodiment, the method is directed to making a host cell in which the lipid-linked oligosaccharide precursors are enriched in $Man_XGlcNAc_2$ structures, where X is 3, 4 or 5 (**Fig. 2**). These structures are transferred in the ER of the host cell onto nascent polypeptide chains by an oligosaccharyl-transferase and may then be processed by treatment with glycosidases (e.g., α-mannosidases) and glycosyltransferases (e.g., GnT1) to produce N-glycans having $GlcNAcMan_XGlcNAc_2$ core structures, wherein X is 3, 4 or 5, and is preferably 3 (**Figs. 2 and 3**). As shown in **Fig. 2**, N-glycans having a $GlcNAcMan_XGlcNAc_2$ core structure where X is greater than 3 may be converted to $GlcNAcMan_3GlcNAc_2$, e.g., by treatment with an α-1,3 and/or α-1,2-1,3 mannosidase activity, where applicable.

[0113] Additional processing of $GlcNAcMan_3GlcNAc_2$ by treatment with glycosyltransferases (e.g., GnTII) produces $GlcNAc_2Man_3GlcNAc_2$ core structures which may then be modified, as desired, e.g., by *ex vivo* treatment or by heterologous expression in the host cell of a set of glycosylation enzymes, including glycosyltransferases, sugar transporters and mannosidases (see below), to become human-like N-glycans. Preferred human-like glycoproteins which may be produced according to the invention include those which comprise *N*-glycans having seven or fewer, or three or fewer, mannose residues; comprise one or more sugars selected from the group consisting of galactose, GlcNAc, sialic acid, and fucose; and comprise at least one oligosaccharide branch comprising the structure NeuNAc-Gal-GlcNAc-Man.

[0114] In one embodiment, the host cell has diminished or depleted Dol-P-Man:$Man_5GlcNAc_2$-PP-Dol Mannosyltransferase activity, which is an activity involved in the first mannosylation step from $Man_5GlcNAc_2$-PP-Dol to

$Man_6GlcNAc_2$-PP-Dol at the luminal side of the ER (e.g., *ALG3* **Fig. 1**; **Fig. 2**). In *S.cerevisiae*, this enzyme is encoded by the *ALG3* gene. As described above, *S.cerevisiae* cells harboring a leaky *alg3-1* mutation accumulate $Man_5GlcNAc_2$-PP-Dol and cells having a deletion in *alg3* appear to transfer $Man_5GlcNAc_2$ structures onto nascent polypeptide chains within the ER. Accordingly, in this embodiment, host cells will accumulate N-glycans enriched in $Man_5GlcNAc_2$ structures which can then be converted to $GlcNAc_2Man_3GlcNAc_2$ by treatment with glycosidases (e.g., with α-1,2 mannosidase, α-1,3 mannosidase or α-1,2-1,3 mannosidase activities (**Fig. 2**).

[0115] As described in **Example 1**, degenerate primers were designed based on an alignment of Alg3 protein sequences from *S. cerevisiae*, *D. melanogaster* and humans (*H. sapiens*) (**Figs. 4 and 5**), and were used to amplify a product from *P. pastoris* genomic DNA. The resulting PCR product was used as a probe to identify and isolate a *P. pastoris* genomic clone comprising an open reading frame (ORF) that encodes a protein having 35% overall sequence identity and 53% sequence similarity to the *S. cerevisiae ALG3* gene (**Figs. 6 and 7**). This *P. pastoris* gene is referred to herein as "*PpALG3*". The *ALG3* gene was similarly identified and isolated from *K. lactis* (**Example 1**; **Figs. 8 and 9**).

[0116] Thus, in another embodiment, the invention provides an isolated nucleic acid molecule having a nucleic acid sequence comprising or consisting of at least forty-five, preferably at least 50, more preferably at least 60 and most preferably 75 or more nucleotide residues of the *P.pastoris ALG 3*gene (**Fig. 6**) and the *K. lactis ALG 3*gene (**Fig. 8**), and homologs, variants and derivatives thereof. The invention also provides nucleic acid molecules that hybridize under stringent conditions to the above-described nucleic acid molecules. Similarly, isolated polypeptides (including muteins, allelic variants, fragments, derivatives, and analogs) encoded by the nucleic acid molecules of the invention are provided (*P.pastoris* and *K. lactis ALG 3*gene products are shown in **Fig. 6 and 8**). In addition, also provided are vectors, including expression vectors, which comprise a nucleic acid molecule of the invention, as described further herein.

[0117] Using gene-specific primers, a construct was made to delete the *PpALG3* gene from the genome of *P. pastoris* (**Example 1**). This strain was used to

generate a host cell depleted in Dol-P-Man:$Man_5GlcNAc_2$-PP-Dol Mannosyltransferase activity and produce lipid-linked $Man_5GlcNAc_2$-PP-Dol precursors which are transferred onto nascent polypeptide chains to produce N-glycans having a $Man_5GlcNAc_2$ carbohydrate structure.

[0118] As described in **Example 2**, such a host cell may be engineered by expression of appropriate mannosidases to produce N-glycans having the desired $Man_3GlcNAc_2$ core carbohydrate structure. Expression of GnTs in the host cell (e.g., by targeting a nucleic acid molecule or a library of nucleic acid molecules as described below) enables the modified host cell to produce N-glycans having one or two GlcNAc structures attached to each arm of the Man3 core structure (i.e., $GlcNAc_1Man_3GlcNAc_2$ or $GlcNAc_2Man_3GlcNAc_2$; see **Fig. 3**). These structures may be processed further using the methods of the invention to produce human-like N-glycans on proteins which enter the secretion pathway of the host cell.

[0119] In another embodiment, the host cell has diminished or depleted dolichyl-P-Man:$Man_6GlcNAc2$-PP-dolichyl α-1,2 mannosyltransferase activity, which is an α-1,2 mannosyltransferase activity involved in the mannosylation step converting $Man_6GlcNAc_2$-PP-Dol to $Man_7GlcNAc_2$-PP-Dol at the luminal side of the ER (see above and **Figs. 1 and 2**). In *S.cerevisiae*, this enzyme is encoded by the *ALG9* gene. Cells harboring an *alg9* mutation accumulate $Man_6GlcNAc_2$-PP-Dol (**Fig. 2**) and transfer $Man_6GlcNAc_2$ structures onto nascent polypeptide chains within the ER. Accordingly, in this embodiment, host cells will accumulate N-glycans enriched in $Man_6GlcNAc_2$ structures which can then be processed down to core Man3 structures by treatment with α-1,2 and α-1,3 mannosidases (see **Fig. 3 and Examples 3 and 4**).

[0120] A host cell in which the *alg9* gene (or gene encoding an equivalent activity) has been deleted is constructed (see, e.g., **Example 3**). Deletion of *ALG9* (or *ALG12*; see below) creates a host cell which produces N-glycans with one or two additional mannoses, respectively, on the 1,6 arm (**Fig. 2**). In order to make the 1,6 core-mannose accessible to N-acetylglucosaminyltransferase II (GnTII) these mannoses have to be removed by glycosidase(s). ER mannosidase typically will remove the terminal 1,2 mannose on the 1,6 arm and subsequently Mannosidase II (alpha 1-3,6 mannosidase) or other mannosidases such as alpha

1,2, alpha1,3 or alpha 1-2,3 mannosidases (e.g., from *Xanthomonas manihotis*; see **Example 4**) can act upon the 1,6 arm and subsequently GnTII can transfer an N-acetylglucosamine, resulting in $GlcNAc_2Man_3$ (**Fig. 2**).

[0121] The resulting host cell, which is depleted for alg9p activity, is engineered to express α-1,2 and α-1,3 mannosidase activity (from one or more enzymes, and preferably, by expression from a nucleic acid molecule introduced into the host cell and which expresses an enzyme targeted to a preferred subcellular compartment (see below). **Example 4** describes the cloning and expression of one such enzyme from *Xanthomonas manihotis*.

[0122] In another embodiment, the host cell has diminished or depleted dolichyl-P-Man:Man7GlcNAc2-PP-dolichyl α-1,6 mannosyltransferase activity, which is an α-1,6 mannosyltransferase activity involved in the mannosylation step converting $Man_7GlcNAc_2$-PP-Dol to $Man_8GlcNAc_2$-PP-Dol (which mannosylates the α-1,6 mannose on the 1,6 arm of the core mannose structure) at the luminal side of the ER (see above and **Figs. 1 and 2**). In *S.cerevisiae*, this enzyme is encoded by the *ALG12* gene. Cells harboring an *alg12* mutation accumulate $Man_7GlcNAc_2$-PP-Dol (**Fig. 2**) and transfer $Man_7GlcNAc_2$ structures onto nascent polypeptide chains within the ER. Accordingly, in this embodiment, host cells will accumulate N-glycans enriched in $Man_7GlcNAc_2$ structures which can then be processed down to core Man3 structures by treatment with α-1,2 and α-1,3 mannosidases (see **Fig. 3 and Examples 3 and 4**).

[0123] As described above for alg9 mutant hosts, the resulting host cell, which is depleted for alg12p activity, is engineered to express α-1,2 and α-1,3 mannosidase activity (e.g., from one or more enzymes, and preferably, by expression from one or more nucleic acid molecules introduced into the host cell and which express an enzyme activity which is targeted to a preferred subcellular compartment (see below).

[0124]

Engineering or Selecting Hosts Optionally Having Decreased Initiating α-1,6 Mannosyltransferase Activity

[0125] In a preferred embodiment, the method of the invention involves making or using a host cell which is both (a) diminished or depleted in the activity of an

alg gene or in one or more activities that mannosylate N-glycans on the α-1,6 arm of the $Man_3GlcNAc_2$ ("Man3") core carbohydrate structure; and (b) diminished or depleted in the activity of an initiating α-1,6-mannosyltransferase, i.e., an initiation specific enzyme that initiates outer chain mannosylation (on the α-1,3 arm of the Man3 cores structure). In *S.cerevisiae*, this enzyme is encoded by the *OCH1* gene. Disruption of the *och1* gene in *S.cerevisiae* results in a phenotype in which N-linked sugars completely lack the poly-mannose outer chain. Previous approaches for obtaining mammalian-type glycosylation in fungal strains have required inactivation of *OCH1* (see, e.g., Chiba, 1998). Disruption of the initiating α-1,6-mannosyltransferase activity in a host cell of the invention is optional, however (depending on the selected host cell), as the Och1p enzyme requires an intact $Man_8GlcNAc$ for efficient mannose outer chain initiation. Thus, the host cells selected or produced according to this invention, which accumulate lipid-linked oligosaccharides having seven or fewer mannose residues will, after transfer, produce hypoglycosylated N-glycans that will likely be poor substrates for Och1p (see, e.g., Nakayama, 1997).

Engineering or Selecting Hosts Having Increased Glucosyltransferase Activity

[0126] As discussed above, glucosylated oligosaccharides are thought to be transferred to nascent polypeptide chains at a much higher rate than their nonglucosylated counterparts. It appears that substrate recognition by the oligosaccharyltransferase complex is enhanced by addition of glucose to the antennae of lipid-linked oligosaccharides. It is thus desirable to create or select host cells capable of optimal glucosylation of the lipid-linked oligosaccharides. In such host cells, underglycosylation will be substantially decreased or even abolished, due to a faster and more efficient transfer of glucosylated Man_5 structures onto the nascent polypeptide chain.

[0127] Accordingly, in another embodiment of the invention, the method is directed to making a host cell in which the lipid-linked N-glycan precursors are transferred efficiently to the nascent polypeptide chain in the ER. In a preferred embodiment, transfer is augmented by increasing the level of glucosylation on the

branches of lipid-linked oligosaccharides which, in turn, will make them better substrates for oligosaccharyltransferase.

[0128] In one preferred embodiment, the invention provides a method for making a human-like glycoprotein which uses a host cell in which one or more enzymes responsible for glucosylation of lipid-linked oligosaccharides in the ER has increased activity. One way to enhance the degree of glucosylation of the lipid-linked oligosaccharides is to overexpress one or more enzymes responsible for the transfer of glucose residues onto the antennae of the lipid-linked oligosaccharide. In particular, increasing α-1,3 glucosyltransferase activity will increase the amount of glucosylated lipid-linked Man_5 structures and will reduce or eliminate the underglycosylation of secreted proteins. In *S.cerevisiae*, this enzyme is encoded by the *ALG6* gene.

[0129] *Saccharomyces cerevisiae ALG6* and its human counterpart have been cloned (Imbach, 1999; Reiss, 1996). Due to the evolutionary conservation of the early steps of glycosylation, *ALG6* loci are expected to be homologous between species and may be cloned based on sequence similarities by anyone skilled in the art. (The same holds true for cloning and identification of *ALG8* and *ALG10* loci from different species.) In addition, different glucosyltransferases from different species can then be tested to identify the ones with optimal activities.

[0130] The introduction of additional copies of an *ALG6* gene and/or the expression of *ALG6* under the control of a strong promoter, such as the GAPDH promoter, is one of several ways to increase the degree of glucosylated lipid-linked oligosaccharides. The *ALG6* gene from *P. pastoris* is cloned and expressed (**Example 5**). *ALG6* nucleic acid and amino acid sequences are show in **Fig. 25** (*S. cerevisiae*) and **Fig. 26** (*P. pastoris*). These sequences are compared to other eukaryotic *ALG6* sequences in **Fig. 27**.

[0131] Accordingly, another embodiment of the invention provides a method to enhance the degree of glucosylation of lipid-linked oligosaccharides comprising the step of increasing alpha-1,3 glucosyltransferase activity in a host cell. The increase in activity may be achieved by overexpression of nucleic acid sequences encoding the activity, e.g., by operatively linking the nucleic acid encoding the activity with one or more heterologous expression control sequences. Preferred

expression control sequences include transcription initiation, termination, promoter and enhancer sequences; RNA splice donor and polyadenylation signals; mRNA stabilizing sequences; ribosome binding sites; protein stabilizing sequences; and protein secretion sequences.

[0132] In another embodiment, the increase in alpha-1,3 glucosyltransferase activity is achieved by introducing a nucleic acid molecule encoding the activity on a multi-copy plasmid, using techniques well known to the skilled worker. In yet another embodiment, the degree of glucosylation of lipid-linked oligosaccharides comprising decreasing the substrate specificity of oligosaccharyl transferase activity in a host cell. This is achieved by, for example, subjecting at least one nucleic acid encoding the activity to a technique such as gene shuffling, *in vitro* mutagenesis, and error-prone polymerase chain reaction, all of which are well-known to one of skill in the art. Naturally, *ALG8* and *ALG10* can be overexpressed in a host cell and tested in a similar fashion.

[0133] Accordingly, in a preferred embodiment, the invention provides a method for making a human-like glycoprotein using a host cell which is engineered or selected so that one or more enzymes responsible for glucosylation of lipid-linked oligosaccharides in the ER has increased activity. In a more preferred embodiment, the invention uses a host cell having both (a) diminished or depleted in the activity of one or more alg gene activities or activities that mannosylate N-glycans on the α-1,6 arm of the $Man_3GlcNAc_2$ ("Man3") core carbohydrate structure and (b) engineered or selected so that one or more enzymes responsible for glucosylation of lipid-linked oligosaccharides in the ER has increased activity. The lipid-linked Man_5 structure found in an *alg3* mutant background, however, is not a preferred substrate for Alg6p. Accordingly, the skilled worker may identify Alg6p, Alg8p and Alg10p with an increased substrate specificity (Gibbs, 2001) e.g., by subjecting nucleic acids encoding such enzymes to one or more rounds of gene shuffling, error prone PCR, or *in vitro* mutagenesis approaches and selecting for increased substrate specificity in a host cell of interest, using molecular biology and genetic selection techniques well known to those of skill in the art. It will be appreciated by the skilled worker that such techniques for improving enzyme substrate specificities in a selected host strain are not limited to this particular

embodiment of the invention but rather, may be used in any embodiment to optimize further the production of human-like N-glycans in a non-human host cell.

[0134] As described, once Man_5 is transferred onto the nascent polypeptide chain, expression of suitable α-1,2-mannosidase(s), as provided by the present invention, will further trim $Man_5GlcNAc_2$ structures to yield the desired core $Man_3GlcNAc_2$ structures. α-1,2-mannosidases remove only terminal α-1,2-linked mannose residues and are expected to recognize the $Man_5GlcNAc_2$ – $Man_7GlcNAc_2$ specific structures made in *alg3, 9* and *12* mutant host cells and in host cells in which homologs to these genes are mutated.

[0135] As schematically presented in **Figure 3**, co-expression of appropriate UDP-sugar-transporter(s) and –transferase(s) will cap the terminal α-1,6 and α-1,3 residues with GlcNAc, resulting in the necessary precursor for mammalian-type complex and hybrid N-glycosylation: $GlcNAc_2Man_5GlcNAc_2$. The peptide-bound N-linked oligosaccharide chain $GlcNAc_2Man_3GlcNAc_2$ (**Figure 3**) then serves as a precursor for further modification to a mammalian-type oligosaccharide structure. Subsequent expression of galactosyl-tranferases and genetically engineering the capacity to transfer sialylic acid will produce a mammalian-type (e.g., human-like) N-glycan structure.

[0136] A desired host cell according to the invention can be engineered one enzyme or more than one enzyme at a time. In addition, a library of genes encoding potentially useful enzymes can be created, and a strain having one or more enzymes with optimal activities or producing the most "human-like" glycoproteins, selected by transforming target host cells with one or more members of the library. Lower eukaryotes that are able to produce glycoproteins having the core *N*-glycan $Man_3GlcNAc_2$ are particularly useful because of the ease of performing genetic manipulations, and safety and efficiency features. In a preferred embodiment, at least one further glycosylation reaction is performed, *ex vivo* or *in vivo,* to produce a human-like N-glycan. In a more preferred embodiment, active forms of glycosylating enzymes are expressed in the endoplasmic reticulum and/or Golgi apparatus of the host cell to produce the desired human-like glycoprotein.

Host Cells

[0137] A preferred non-human host cell of the invention is a lower eukaryotic cell, e.g., a unicellular or filamentous fungus, which is diminished or depleted in the activity of one or more *alg* gene activities (including an enzymatic activity which is a homolog or equivalent to an *alg* activity). Another preferred host cell of the invention is diminished or depleted in the activity of one or more enzymes (other than alg activities) that mannosylate the α-1,6 arm of a lipid-linked oligosaccharide structure.

[0138] While lower eukaryotic host cells are preferred, a wide variety of host cells having the aforementioned properties are envisioned as being useful in the methods of the invention. Plant cells, for instance, may be engineered to express a human-like glycoprotein according to the invention. Likewise, a variety of non-human, mammalian host cells may be altered to express more human-like glycoproteins using the methods of the invention. An appropriate host cell can be engineered, or one of the many such mutants already described in yeasts may be used. A preferred host cell of the invention, as exemplified herein, is a hypermannosylation-minus (*OCH1*) mutant in *Pichia pastoris* which has further been modified to delete the *alg3* gene. Other preferred hosts are *Pichia pastoris* mutants having *och1* and *alg 9* or *alg12* mutations.

Formation of complex N-glycans

[0139] The sequential addition of sugars to the modified, nascent N-glycan structure involves the successful targeting of glucosyltransferases into the Golgi apparatus and their successful expression. This process requires the functional expression, e.g., of GnT I, in the early or medial Golgi apparatus as well as ensuring a sufficient supply of UDP-GlcNAc (e.g., by expression of a UDP-GlcNAc transporter).

[0140] To characterize the glycoproteins and to confirm the desired glycosylation, the glycoproteins were purified, the N-glycans were PNGase-F released and then analyzed by MALDI-TOF-MS (**Example 2**). Kringle 3 domain

of human plasminogen was used as the reporter protein. This soluble glycoprotein was produced in *P. pastoris* in an *alg3*, *och1* knockout background (**Example 2**).

[0141] $GlcNAcMan_5GlcNAc_2$ was produced as the predominant N-glycan after addition of human GnT I, and *K. lactis* UDP-GlcNAc transporter in **Fig. 16** (**Example 2**). The mass of this N-glycan is consistent with the mass of $GlcNAcMan_5GlcNAc_2$ at 1463 (m/z). To confirm the addition of the GlcNAc onto $Man_5GlcNAc_2$, a β-N-hexosaminidase digest was performed, which revealed a peak at 1260 (m/z), consistent with the mass of $Man_5GlcNAc_2$ (**Fig.17**).

[0142] The N-glycans from the *alg3 och1* deletion in one strain PBP3 (Example 2) provided two distinct peaks at 1138 (m/z) and 1300 (m/z), which is consistent with structures $GlcNAcMan_3GlcNAc_2$ and $GlcNAcMan_4GlcNAc_2$ (**Fig. 18**). After an *in vitro* α1,2-mannosidase digestion for redundant mannoses, a peak eluted at 1138 (m/z), which is consistent with $GlcNAcMan_3GlcNAc_2$ (**Fig. 19**). To confirm the addition of the GlcNAc onto the $Man_3GlcNAc_2$ structure, a β-N-hexosaminidase digest was performed, which revealed a peak at 934 (m/z), consistent with the mass of $Man_3GlcNAc_2$ (**Fig. 20**).

[0143] The addition of the second GlcNAc onto $GlcNAcMan_3GlcNAc_2$ is shown in **Fig. 21**. The peak at 1357 (m/z) corresponds to $GlcNAc_2Man_3GlcNAc_2$. To confirm the addition of the two GlcNAcs onto the core mannose structure $Man_3GlcNAc_2$, another β-N-hexosaminidase digest was performed, which revealed a peak at 934 (m/z), consistent with the mass of $Man_3GlcNAc_2$ (**Fig. 22**). This is conclusive data displaying a complex-type glycoprotein made in yeast cells.

[0144] The *in vitro* addition of UDP-galactose and β 1,4-galactosyltransferase onto the $GlcNAc_2Man_3GlcNAc_2$ resulted in a peak at 1664 (m/z), which is consistent with the mass of $Gal_2GlcNAc_2Man_3GlcNAc_2$ (**Fig. 23**) Finally, the *in vitro* addition of CMP-N-acetylneuraminic acid and sialyltransferase resulted in a peak at 2248 (m/z), which is consistent with the mass of $NANA_2Gal_2GlcNAc_2Man_3GlcNAc_2$ (**Fig. 24**). The above data supports the use of non-mammalian host cells, which are capable of producing complex human-like glycoproteins.

Targeting of glycosyl- and galactosyl-transferases to specific organelles.

[0145] Much work has been dedicated to revealing the exact mechanism by which these enzymes are retained and anchored to their respective organelle. Although complex, evidence suggests that, stem region, membrane spanning region and cytoplasmic tail individually or in concert direct enzymes to the membrane of individual organelles and thereby localize the associated catalytic domain to that locus.

[0146] The method by which active glycosyltransferases can be expressed and directed to the appropriate organelle such that a sequential order of reactions may occur, that leads to complex N-glycan formation, is as follows:

(A) Establish a DNA library of regions that are known to encode proteins/peptides that mediate localization to a particular location in the secretory pathway (ER, Golgi and trans Golgi network). A limited selection of such enzymes and their respective location is shown in Table 1. These sequences may be selected from the host to be engineered as well as other related or unrelated organism. Generally such sequences fall into three categories: (1) N-terminal sequences encoding a cytosolic tail (ct), a transmembrane domain (tmd) and part of a somewhat more ambiguously defined stem region (sr), which together or individually anchor proteins to the inner (lumenal) membrane of the Golgi, (2) retrieval signals which are generally found at the C-terminus such as the HDEL or KDEL tetrapeptide, and (3) membrane spanning nucleotide sugar transporters, which are known to locate in the Golgi. In the first case, where the localization region consists of various elements (ct, tmd and sr) the library is designed such that the ct, the tmd and various parts of the stem region are represented. This may be accomplished by using PCR primers that bind to the 5' end of the DNA encoding the cytosolic region and employing a series of opposing primers that bind to various parts of the stem region. In addition one would create fusion protein constructs that encode sugar nucleotide transporters and known retrieval signals.

(B) A second step involves the creation of a series of fusion protein constructs, that encode the above mentioned localization sequences and the catalytic domain of a particular glycosyltransferase cloned in frame to such localization sequence (e.g. GnT I, GalT, Fucosyltransferase or ST). In the case of a sugar nucleotide

transporter fused to a catalytic domain one may design such constructs such that the catalytic domain (e.g. GnT I) is either at the N- or the C-terminus of the resulting polypeptide. The catalytic domain, like the localization sequence, may be derived from various different sources. The choice of such a catalytic domains may be guided by the knowledge of the particular environment in which the catalytic domain is to be active. For example, if a particular glycosyltransferase is to be active in the late Golgi, and all known enzymes of the host organism in the late Golgi have a pH optimum of 7.0, or the late Golgi is known to have a particular pH, one would try to select a catalytic domain that has maximum activity at that pH. Existing *in vivo* data on the activity of such enzymes, in particular hosts, may also be of use. For example, Schwientek and coworkers showed that GalT activity can be engineered into the Golgi of *S.cerevisiae* and showed that such activity was present by demonstrating the transfer of some Gal to existing $GlcNAc_2$ in an *alg* mutant of *S. cerevisiae*. In addition, one may perform several rounds of gene shuffling or error prone PCR to obtain a larger diversity within the pool of fusion constructs, since it has been shown that single amino mutations may drastically alter the activity of glycoprotein processing enzymes (Romero et al., 2000). Full length sequences of glycosyltransferases and their endogenous anchoring sequence may also be used. In a preferred embodiment, such localization/catalytic domain libraries are designed to incorporate existing information on the sequential nature of glycosylation reactions in higher eukaryotes. In other words, reactions known to occur early in the course of glycoprotein processing require the targeting of enzymes that catalyze such reactions to an early part of the Golgi or the ER. For example, the trimming of $Man_8GlcNAc_2$ to $Man_5GlcNAc_2$ is an early step in complex N-glycan formation. Since protein processing is initiated in the ER and then proceeds through the early, medial and late Golgi, it is desirable to have this reaction occur in the ER or early Golgi. When designing a library for mannosidase I localization, one thus attempts to match ER and early Golgi targeting signals with the catalytic domain of mannosidase I.

[0147] Upon transformation of the host strain with the fusion construct library a selection process is used to identify which particular combination of localization

sequence and catalytic domain in fact have the maximum effect on the carbohydrate structure found in such host strain. Such selection can be based on any number of assays or detection methods. They may be carried out manually or may be automated through the use of high troughput screening equipment.

[0148] In another example, GnT I activity is required for the maturation of complex N-glycans, because only after addition of GlcNAc to the terminal α1,3 mannose residue may further trimming of such a structure to the subsequent intermediate $GlcNAcMan_3GlcNAc_2$ structure occur. Mannosidase II is most likely not capable of removing the terminal α1,3- and α1,6- mannose residues in the absence of a terminal β1,2-GlcNAc and thus the formation of complex N-glycans will not proceed in the absence of GnT I activity (Schachter, 1991). Alternatively, one may first engineer or select a strain that makes sufficient quantities of $Man_5GlcNAc_2$ as described in this invention by engineering or selecting a strain deficient in Alg3P activity. In the presence of sufficient UDP-GlcNAc transporter activity, as may be achieved by engineering or selecting a strain that has such UDP-GlcNAc transporter activity, GlcNAc can be added to the terminal α-1,3 residue by GnTI as *in vitro* a Man_3 structure is recognized by by rat liver GnTI (Moller, 1992).

[0149] In another approach, one may incorporate the expression of a UDP-GlcNAc transporter into the library mentioned above such that the desired construct will contain: (1) a region by which the transformed construct is maintained in the cell (e.g. origin of replication or a region that mediates chromosomal integration), (2) a marker gene that allows for the selection of cells that have been transformed, including counterselectable and recyclable markers such as *ura3* or *T-urf13* (Soderholm, 2001) or other well characterized selection-markers (e.g, *his4, bla, Sh ble* etc.), (3) a gene encoding a UDP-GlcNAc transporter (e.g. from *K.lactis*, (Abeijon, 1996), or from *H.sapiens* (Ishida, 1996), and (4) a promotor activating the expression of the above mentioned localization/catalytic domain fusion construct library.

[0150] After transformation of the host with the library of fusion constructs described above, one may screen for those cells that have the highest concentration of terminal GlcNAc on the cell surface, or secrete the protein with the highest

terminal GlcNAc content. Such a screen may be based on a visual method, like a staining procedure, the ability to bind specific terminal GlcNAc binding antibodies or lectins conjugated to a marker (such lectins are available from E.Y. Laboratories Inc., San Mateo, CA), the reduced ability of specific lectins to bind to terminal mannose residues, the ability to incorporate a radioactively labeled sugar *in vitro*, altered binding to dyes or charged surfaces, or may be accomplished by using a Fluorescence Assisted Cell Sorting (FACS) device in conjunction with a fluorophore labeled lectin or antibody (Guillen, 1998). It may be advantageous to enrich particular phenotypes within the transformed population with cytotoxic lectins. U.S. Patent No. 5,595,900 teaches several methods by which cells with a desired extra-cellular carbohydrate structures may be identified. Repeatedly carrying out this strategy allows for the sequential engineering of more and more complex glycans in lower eukaryotes.

[0151] After transformation, one may select for transformants that allow for the most efficient transfer of GlcNAc by GlcNAc Transferase II from UDP-GlcNAc in an *in vitro* assay. This screen may be carried out by growing cells harboring the transformed library under selective pressure on an agar plate and transferring individual colonies into a 96-well microtiter plate. After growing the cells, the cells are centrifuged, the cells resuspended in buffer, and after addition of UDP-GlcNAc and GnT V, the release of UDP is determined either by HPLC or an enzyme linked assay for UDP. Alternatively, one may use radioactively labeled UDP-GlcNAc and GnT V, wash the cells and then look for the release of radioactive GlcNAc by N-actylglucosaminidase. All this may be carried manually or automated through the use of high throughput screening equipment.

[0152] Transformants that release more UDP, in the first assay, or more radioactively labeled GlcNAc in the second assay, are expected to have a higher degree of $GlcNAcMan_3GlcNAc_2$ (**Fig. 3**) on their surface and thus constitute the desired phenotype. Alternatively, one may any use any other suitable screen such as a lectin binding assay that is able to reveal altered glycosylation patterns on the surface of transformed cells. In this case the reduced binding of lectins specific to terminal mannoses may be a suitable selection tool. *Galantus nivalis* lectin binds specifically to terminal α-1,3 mannose, which is expected to be reduced if

sufficient mannosedase II activity is present in the Golgi. One may also enrich for desired transformants by carrying out a chromatographic separation step that allows for the removal of cells containing a high terminal mannose content. This separation step would be carried out with a lectin column that specifically binds cells with a high terminal mannose content (e.g *Galantus nivalis* lectin bound to agarose , Sigma, St.Louis, MO) over those that have a low terminal mannose content. In addition, one may directly create such fusion protein constructs, as additional information on the localization of active carbohydrate modifying enzymes in different lower eukaryotic hosts becomes available in the scientific literature. For example, the prior art teaches us that human β1,4-GalTr can be fused to the membrane domain of MNT, a mannosyltransferase from *S. cerevisiae*, and localized to the Golgi apparatus while retaining its catalytic activity (Schwientek et al., 1995). If *S. cerevisiae* or a related organism is the host to be engineered one may directly incorporate such findings into the overall strategy to obtain complex N-glycans from such a host. Several such gene fragments in *P.pastoris* have been identified that are related to glycosyltransferases in S.cerevisiae and thus could be used for that purpose.

Table 1

Gene or sequence	Organism	Function	Location of gene product
MnsI	*S.cerevisiae*	mannosidase	ER
Och1	*S.cerevisiae*	1,6-mannosyltransferase	Golgi (cis)
Mnn2	*S.cerevisiae*	1,2-mannosyltransferase	Golgi (medial)
Mnn1	*S.cerevisiae*	1,3-mannosyltransferase	Golgi (trans)
Och1	*P.pastoris*	1,6-mannosyltransferase	Golgi (cis)
2,6 ST	*H.sapiens* *S. frugiperda*	2,6-sialyltransferase	trans-Golgi network
β1,4 Gal T	*bovine milk*	UDP-Gal transporter	Golgi
Mnt1	*S.cerevisiae*	1,2-mannosyltransferase	Golgi (cis)
HDEL at C-terminus	*S.cerevisiae*	retrieval signal	ER

Integration Sites

[0153] As one ultimate goal of this genetic engineering effort is a robust protein production strain that is able to perform well in an industrial fermentation process, the integration of multiple genes into the host (e.g., fungal) chromosome involves careful planning. The engineered strain will most likely have to be transformed with a range of different genes, and these genes will have to be transformed in a stable fashion to ensure that the desired activity is maintained throughout the fermentation process. Any combination of the following enzyme activities will have to be engineered into the fungal protein expression host: sialyltransferases, mannosidases, fucosyltransferases, galactosyltransferases, glucosyltransferases, GlcNAc transferases, ER and Golgi specific transporters (e.g. syn and antiport transporters for UDP-galactose and other precursors), other enzymes involved in the processing of oligosaccharides, and enzymes involved in the synthesis of activated oligosaccharide precursors such as UDP-galactose, CMP-N-acetylneuraminic acid. At the same time, a number of genes which encode enzymes known to be characteristic of non-human glycosylation reactions, will have to be deleted. Such genes and their corresponding proteins have been extensively characterized in a number of lower eukaryotes (e.g. *S.cerevisiae, T.reesei, A. nidulans* etc.), thereby providing a list of known glycosyltransferases in lower eukaryotes, their activities and their respective genetic sequence. These genes are likely to be selected from the group of mannosyltransferases e.g. 1,3 mannosyltransferases (e.g. MNN1 in *S.cerevisiae*) (Graham, 1991), 1,2 mannosyltransferases (e.g. KTR/KRE family from *S.cerevisiae*), 1,6 mannosyltransferases (OCH1 from *S.cerevisiae*), mannosylphosphate transferases (MNN4 and MNN6 from *S.cerevisiae*) and additional enzymes that are involved in aberrant i.e. non human glycosylation reactions. Many of these genes have in fact been deleted individually giving rise to viable phenotypes with altered glycosylation profiles. Examples are shown in Table 2:

Table 2.

Strain	Mutant	Structure wild type	Structure mutant	Authors
Schizosaccharomyces pombe	OCH1	Mannan (i.e. $Man_{>9}GlcNAc_2$)	$Man_8GlcNAc_2$	Yoko-o et al., 2001

S.cerevisiae	*OCH1, MNN1*	Mannan (i.e. $Man_{>9}GlcNAc_2$)	$Man_8GlcNAc_2$	Nakanishi-Shindo et al,. 1993
S.cerevisiae	*OCH1, MNN1, MNN4*	Mannan (i.e. $Man_{>9}GlcNAc_2$)	$Man_8GlcNAc_2$	Chiba et al., 1998

As any strategy to engineer the formation of complex N-glycans into a lower eukaryote involves both the elimination as well as the addition of glycosyltransferase activities, a comprehensive scheme will attempt to coordinate both requirements. Genes that encode enzymes that are undesirable serve as potential integration sites for genes that are desirable. For example, 1,6 mannosyltransferase activity is a hallmark of glycosylation in many known lower eukaryotes. The gene encoding alpha-1,6 mannosyltransferase (*OCH1*) has been cloned from *S.cerevisiae* and mutations in the gene give raise to a viable phenotype with reduced mannosylation. The gene locus encoding alpha-1,6 mannosyltransferase activity therefor is a prime target for the integration of genes encoding glycosyltransferase activity. In a similar manner, one can choose a range of other chromosomal integration sites that, based on a gene disruption event in that locus, are expected to: (1) improve the cells ability to glycosylate in a more human like fashion, (2) improve the cells ability to secrete proteins, (3) reduce proteolysis of foreign proteins and (4) improve other characteristics of the process that facilitate purification or the fermentation process itself.

Providing sugar nucleotide precursors

[0154] A hallmark of higher eukaryotic glycosylation is the presence of galactose, fucose, and a high degree of terminal sialic acid on glycoproteins. These sugars are not generally found on glycoproteins produced in yeast and filamentous fungi and the method discussed above allows for the engineering of strains that localize glycosyltransferase in the desired organelle. Formation of complex N-glycan synthesis is a sequential process by which specific sugar residues are removed and attached to the core oligosaccharide structure. In higher eukaryotes, this is achieved by having the substrate sequentially exposed to various processing enzymes. These enzymes carry out specific reactions depending on their particular location within the entire processing cascade. This "assembly line"

consists of ER, early, medial and late Golgi, and the trans Golgi network all with their specific processing environment. To recreate the processing of human glycoproteins in the Golgi and ER of lower eukaryotes, numerous enzymes (e.g. glycosyltransferases, glycosidases, phosphatases and transporters) have to be expressed and specifically targeted to these organelles, and preferably, in a location so that they function most efficiently in relation to their environment as well as to other enzymes in the pathway. **[0155]** Several individual glycosyltransferases have been cloned and expressed in *S.cerevisiae* (GalT, GnT I), *Aspergillus nidulans* (GnT I) and other fungi, without however demonstrating the desired outcome of "humanization" on the glycosylation pattern of the organisms (Yoshida, 1995; Schwientek, 1995; Kalsner, 1995). It was speculated that the carbohydrate structure required to accept sugars by the action of such glycosyltransferases was not present in sufficient amounts. While this most likely contributed to the lack of complex N-glycan formation, there are currently no reports of a fungus supplying a $Man_5GlcNAc_2$ structure, having GnT I activity and having UDP-Gn transporter activity engineered into the fungus. It is the combination of these three biochemical events that are required for hybrid and complex N-glycan formation.

[0156] In humans, the full range of nucleotide sugar precursors (e.g. UDP-N-acetylglucosamine, UDP-N-acetylgalactosamine, CMP-N-acetylneuraminic acid, UDP-galactose, etc.) are generally synthesized in the cytosol and transported into the Golgi, where they are attached to the core oligosaccharide by glycosyltransferases. To replicate this process in lower eukaryotes, sugar nucleoside specific transporters have to be expressed in the Golgi to ensure adequate levels of nucleoside sugar precursors (Sommers, 1981; Sommers, 1982; Perez, 1987). A side product of this reaction is either a nucleoside diphosphate or monophosphate. While monophosphates can be directly exported in exchange for nucleoside triphosphate sugars by an antiport mechanism, diphospho nucleosides (e.g. GDP) have to be cleaved by phosphatases (e.g. GDPase) to yield nucleoside monophosphates and inorganic phosphate prior to being exported. This reaction appears to be important for efficient glycosylation, as GDPase from *S.cerevisiae* has been found to be necessary for mannosylation. However, the enzyme only has

10% of the activity towards UDP (Berninsone, 1994). Lower eukaryotes often do not have UDP specific diphosphatase activity in the Golgi since they do not utilize UDP-sugar precursors for glycoprotein synthesis in the Golgi.

[0157] *Schizosaccharomyces pombe*, a yeast found to add galactose residues to cell wall polysaccharides (from UDP-galactose) was found to have specific UDPase activity further suggesting the requirement for such an enzyme (Berninsone et al., 1994). UDP is known to be a potent inhibitor of glycosyltransferases and the removal of this glycosylation side product is important in order to prevent glycosyltransferase inhibition in the lumen of the Golgi (Khatara et al., 1974). Thus, one may need to provide for the removal of UDP, which is expected to accumulate in the Golgi of such an engineered strains (Berninsone, 1995; Beaudet, 1998). **[0158]** In another example, 2,3 sialyltransferase and 2,6 sialyltransferase cap galactose residues with sialic acid in the trans-Golgi and TGN of humans leading to a mature form of the glycoprotein. To reengineer this processing step into a metabolically engineered yeast or fungus will require (1) 2,3-sialyltransferase activity and (2) a sufficient supply of CMP-N-acetyl neuraminic acid, in the late Golgi of yeast. To obtain sufficient 2,3-sialyltransferase activity in the late Golgi, the catalytic domain of a known sialyltransferase (e.g. from humans) has to be directed to the late Golgi in fungi (see above). Likewise, transporters have to be engineered to that allow the transport of CMP-N-acetyl neuraminic acid into the late Golgi. There is currently no indication that fungi synthesize sufficient amounts of CMP-N-acetyl neuraminic acid, not to mention the transport of such a sugar-nucleotide into the Golgi. Consequently, to ensure the adequate supply of substrate for the corresponding glycosyltransferases, one has to metabolically engineer the production of CMP-sialic acid into the fungus.

Methods for providing sugar nucleotide precursors to the Golgi apparatus:

UDP-N-acetyl-glucosamine

[0159] The cDNA of human UDP-N-acetylglucosamine transporter, which was recognized through a homology search in the expressed sequence tags database (dbEST) was cloned by Ishida and coworkers (Ishida, 1999). Guillen and

coworkers have cloned the mammalian Golgi membrane transporter for UDP-N-acetylglucosamine by phenotypic correction with cDNA from canine kidney cells (MDCK) of a recently characterized *Kluyveromyces lactis* mutant deficient in Golgi transport of the above nucleotide sugar (Guillen, 1998). Their results demonstrate that the mammalian Golgi UDP-GlcNAc transporter gene has all of the necessary information for the protein to be expressed and targeted functionally to the Golgi apparatus of yeast and that two proteins with very different amino acid sequences may transport the same solute within the same Golgi membrane (Guillen, 1998).

GDP-Fucose

[0160] The rat liver Golgi membrane GDP-fucose transporter has been identified and purified by Puglielli, L. and C. B. Hirschberg (Puglielli, 1999). The corresponding gene has not been identified however N-terminal sequencing can be used for the design of oligonucleotide probes specific for the corresponding gene. These oligonucleotides can be used as probes to clone the gene encoding for GDP-fucose transporter.

UDP-Galactose

[0161] Two heterologous genes, gmal2(+) encoding alpha 1,2-galactosyltransferase (alpha 1,2 GalT) from *Schizosaccharomyces pombe* and (hUGT2) encoding human UDP-galactose (UDP-Gal) transporter, have been functionally expressed in *S.cerevisiae* to examine the intracellular conditions required for galactosylation. Correlation between protein galactosylation and UDP-galactose transport activity indicated that an exogenous supply of UDP-Gal transporter, rather than alpha 1,2 GalT played a key role for efficient galactosylation in *S.cerevisiae* (Kainuma, 1999). Likewise a UDP-galactose transporter from *S. pombe* was cloned (Aoki, 1999; Segawa, 1999).

CMP-N-acetylneuraminic acid (CMP-Sialic acid)

[0162] Human CMP-sialic acid transporter (hCST) has been cloned and expressed in Lec 8 CHO cells (Aoki, 1999; Eckhardt, 1997). The functional expression of the murine CMP-sialic acid transporter was achieved in *Saccharomyces cerevisiae* (Berninsone, 1997). Sialic acid has been found in some fungi, however it is not clear whether the chosen host system will be able to supply

sufficient levels of CMP-Sialic acid. Sialic acid can be either supplied in the medium or alternatively fungal pathways involved in sialic acid synthesis can also be integrated into the host genome.

Diphosphatases

[0163] When sugars are transferred onto a glycoprotein, either a nucleoside diphosphate or monophosphate, is released from the sugar nucleotide precursors. While monophosphates can be directly exported in exchange for nucleoside triphosphate sugars by an antiport mechanism, diphospho nucleosides (e.g. GDP) have to be cleaved by phosphatases (e.g. GDPase) to yield nucleoside monophosphates and inorganic phosphate prior to being exported. This reaction appears to be important for efficient glycosylation, as GDPase from *S.cerevisiae* has been found to be necessary for mannosylation. However, the enzyme only has 10% of the activity towards UDP (Berninsone, 1994). Lower eukayotes often do not have UDP specific diphosphatase activity in the Golgi since they do not utilize UDP-sugar precursors for glycoprotein synthesis in the Golgi. *Schizosaccharomyces pombe*, a yeast found to add galactose residues to cell wall polysaccharides (from UDP-galactose) was found to have specific UDPase activity further suggesting the requirement for such an enzyme (Berninsone, 1994). UDP is known to be a potent inhibitor of glycosyltransferases and the removal of this glycosylation side product is important in order to prevent glycosyltransferase inhibition in the lumen of the Golgi (Khatara et al. 1974).

Expression Of GnTs To Produce Complex N-glycans

Expression Of GnT-III To Boost Antibody Functionality

[0164] The addition of an N-acetylglucosamine to the $GlcNAc_1Man_3GlcNAc_2$ structure by N-acetylglucosaminyltransferases II and III yields a so-called bisected N-glycan $GlcNAc_3Man_3GlcNAc_2$ **(Fig. 3)**. This structure has been implicated in greater antibody-dependent cellular cytotoxicity (ADCC) (Umana et al. 1999). Re-engineering glycoforms of immunoglobulins expressed by mammalian cells is a tedious and cumbersome task. Especially in the case of GnTIII, where over-

expression of this enzyme has been implicated in growth inhibition, methods involving regulated (inducible) gene expression had to be employed to produce immunoglobulins with bisected N-glycans (Umana et al 1999a, 1999b).

[0165] Accordingly, in another embodiment, the invention provides systems and methods for producing human-like N-glycans having bisecting N-acetylglucosamine (GlcNAcs) on the core mannose structure. In a preferred embodiment, the invention provides a system and method for producing immunoglobulins having bisected N-glycans. The systems and methods described herein will not suffer from previous problems, e.g., cytotoxicity associated with overexpression of GnTIII or ADCC, as the host cells of the invention are engineered and selected to be viable and preferably robust cells which produce N-glycans having substantially modified human-type glycoforms such as $GlcNAc_2Man_3GlcNAc_2$. Thus, addition of a bisecting N-acetylglucosamine in a host cell of the invention will have a negligible effect on the growth-phenotype or viability of those host cells.

[0166] In addition, previous work (Umana) has shown that there is no linear correlation between GnTIII expression levels and the degree of ADCC. Finding the optimal expression level in mammalian cells and maintaining it throughout an FDA approved fermentation process seems to be a challenge. However, in cells of the invention, such as fungal cells, finding a promoter of appropriate strength to establish a robust, reliable and optimal GnTIII expression level is a comparatively easy task for one of skill in the art.

[0167] A host cell such as a yeast strain capable of producing glycoproteins with bisecting N-glycans is engineered according to the invention, by introducing into the host cell a GnTIII activity (**Example 6**). Preferably, the host cell is transformed with a nucleic acid that encodes GnTIII (see, e.g., **Fig. 32**) or a domain thereof having enzymatic activity, optionally fused to a heterologous cell signal targeting peptide (e.g., using the libraries and associated methods of the invention.) Host cells engineereded to express GnTIII will produce higher antibody titers than mammalian cells are capable of. They will also produce antibodies with higher potency with respect to ADCC.

[0168] Antibodies produced by mammalian cell lines transfected with GnTIII have been shown to be as effective as antibodies produced by non-transfected cell-lines, but at a 10-20 fold lower concentration (Davies et al. 2001). An increase of productivity of the production vehicle of the invention over mammalian systems by a factor of twenty, and a ten-fold increase of potency will result in a net-productivity improvement of two hundred. The invention thus provides a system and method for producing high titers of an antibody having high potency (e.g., up to several orders of magnitude more potent than what can currently be produced). The system and method is safe and provides high potency antibodies at low cost in short periods of time. Host cells engineered to express GnT III according to the invention produce immunoglobulins having bisected N-glycans at rates of at least 50 mg/liter/day to at least 500 mg/liter/day. In addition, each immunoglobulin (Ig) molecule (comprising bisecting GlcNAcs) is more potent than the same Ig molecule produced without bisecting GlcNAcs.

Cloning and expression of GnT-IV and GnT-V

[0169] All branching structures in complex N-glycans are synthesized on a common core-pentasaccharide ($Man_3GlcNAc_2$ or Man alpha1-6(Man alpha1-3)Man beta1-4 GlcNAc beta1-4 GlcNAc beta1-4 or $Man_3GlcNAc_2$) by N-acetylglucosamine transferases (GnTs) -I to -VI (Schachter H et al. (1989) *Methods Enzymo*;179:351-97). Current understanding of the biosynthesis of more highly branched N-glycans suggests that after the action of GnTII (generation of $GlcNAc_2Man_3GlcNAc_2$ structures) GnTIV transfers GlcNAc from UDP-GlcNAc in beta1,4 linkage to the Man alpha1,3 Man beta1,4 arm of $GlcNAc_2Man_3GlcNAc_2$ N-glycans (Allen SD et al. (1984) *J Biol Chem.* Jun 10;259(11):6984-90; and Gleeson PA and Schachter H.J (1983); *J.Biol Chem* 25;258(10):6162-73) resulting in a triantennary agalacto sugar chain. This N-glycan (GlcNAc beta1-2 Man alpha1-6(GlcNAc beta1-2 Man alpha1-3) Man beta1-4 GlcNAc beta 1-4 GlcNAc beta1,4 Asn) is a common substrate for GnT-III and –V, leading to the synthesis of bisected, tri-and tetra-antennary structures. Where the action of GnTIII results in a bisected N-glycan and where GnTV catalyzes the addition of beta 1-6GlcNAc to the alpha 1-6 mannosyl core, creating the beta 1-6 branch. Addition of galactose

and sialic acid to these branches leads to the generation of a fully sialylated complex N-glycan.

[0170] Branched complex N-glycans have been implicated in the physiological activity of therapeutic proteins, such as human erythropoietin (hEPO). Human EPO having bi-antennary structures has been shown to have a low activity, whereas hEPO having tetra-antennary structures resulted in slower clearance from the bloodstream and thus in higher activity (Misaizu T et al. (1995) *Blood* Dec 1;86(11):4097-104).

[0171] With DNA sequence information, the skilled worker can clone DNA molecules encoding GnT IV and/or V activities (**Example 6; Figs. 33 and 34**). Using standard techniques well-known to those of skill in the art, nucleic acid molecules encoding GnT IV or V (or encoding catalytically active fragments thereof) may be inserted into appropriate expression vectors under the transcriptional control of promoters and other expression control sequences capable of driving transcription in a selected host cell of the invention, e.g., a fungal host such as *Pichia sp.*, *Kluyveromyces sp.* and *Aspergillus sp.*, as described herein, such that one or more of these mammalian GnT enzymes may be actively expressed in a host cell of choice for production of a human-like complex glycoprotein.

[0172] The following are examples which illustrate the compositions and methods of this invention. These examples should not be construed as limiting: the examples are included for the purposes of illustration only.

EXAMPLE 1

Identification, cloning and deletion of the *ALG3* gene in *P.pastoris* and *K.lactis*.

[0173] Degenerate primers were generated based on an alignment of Alg3 protein sequences from *S. cerevisiae*, *H. sapiens*, and *D. melanogaster* and were used to amplify an 83 bp product from *P. pastoris* genomic DNA: 5'-GGTGTTTTGTTTTCTAGATCTTTGCAYTAYCARTT-3' and 5'-AGAATTTGGTGGGTAAGAATTCCARCACCAYTCRTG-3' The resulting PCR product was cloned into the pCR2.1 vector (Invitrogen, Carlsbad, CA) and

seqence analysis revealed homology to known *ALG3/RHK1/NOT56* homologs (Genbank NC_001134.2, AF309689, NC_003424.1). Subsequently, 1929 bp upstream and 2738 bp downstream of the initial PCR product were amplified from a *P. pastoris* genomic DNA library (Boehm, T. Yeast 1999 May;15(7):563-72) using the internal oligonucleotides 5'- CCTAAGCTGGTATGCGTTCTCTTTGCCATATC-3' and 5'-GCGGCATAAACAATAATAGATGCTATAAAG-3' along with T3 (5'-AATTAACCCTCACTAAAGGG-3') and T7 (5'-GTAA TACGACTCACTATAGGGC-3') (Integrated DNA Technologies, Coralville, IA) in the backbone of the library bearing plasmid lambda ZAP II (Stratagene, La Jolla, CA). The resulting fragments were cloned into the pCR2.1-TOPO vector (Invitrogen) and sequenced. From this sequence, a 1395 bp ORF was identified that encodes a protein with 35% identity and 53% similarity to the *S. cerevisiae ALG3* gene (using BLAST programs). The gene was named *PpALG3*.

[0174] The sequence of *PpALG3*was used to create a set of primers to generate a deletion construct of the *PpALG3* gene by PCR overlap (Davidson et al, 2002 Microbiol. 148(Pt 8):2607-15). Primers below were used to amplify 1 kb regions 5' and 3' of the *PpALG3* ORF and the KAN^R gene, respectively:
RCD142 (5'-CCACATCATCCGTGCTACATATAG-3'),
RCD144 (5'-ACGAGGCAAGCTAAACAGATCTCGAAGTATCGAGGGTTAT CCAG-3'),
RCD145 (5'-CCATCCAGTGTCGAAAACGAGCCAATGGTTCATGTCTATA AATC-3'),
RCD147 (5'-AGCCTCAGCGCCAACAAGCGATGG-3'),
RCD143 (5'-CTGGATAACCCTCGATACTTCGAGATCTGTTTAGCTTGCC TCGT-3'), and
RCD146 (5'-GATTTATAGACATGAACCATTGGCTCGTTTTCGACACTGG ATGG-3').
Subsequently, primers RCD142 and RCD147 were used to overlap the three resulting PCR products into a single 3.6 kb *alg3::KAN*^{R} deletion allele.

Identification, cloning and deletion of the *ALG3* gene in *K.lactis*.

[0175] The *ALG3p* sequences from *S. cerevisiae, Drosophila melanogaster, Homo sapiens etc* were aligned with *K. lactis* sequences (PENDANT EST database). Regions of high homology that were in common homologs but distinct in exact sequence from the homologs were used to create pairs of degenerate primers that were directed against genomic DNA from the *K. lactis* strain MG1/2 (Bianchi et al, 1987). In the case of *ALG3,* PCR amplification with primers KAL-1 (5'-ATCCTTTACCGATGCTGTAT-3') and KAL-2 (5'-ATAACAGTATGTGTTACACGCGTGTAG-3') resulted in a product that was cloned and sequenced and the predicted translation was shown to have a high degree of homology to Alg3p proteins (>50% to *S. cerevisiae* Alg3p).

[0176] The PCR product was used to probe a Southern blot of genomic DNA from *K. lactis* strain (MG1/2) with high stringency (Sambrook et al, 1989). Hybridization was observed in a pattern consistent with a single gene. This Southern blot was used to map the genomic loci. Genomic fragments were cloned by digesting genomic DNA and ligating those fragments in the appropriate size-range into pUC19 to create a *K. lactis* subgenomic library. This subgenomic library was transformed into *E. coli* and several hundred clones were tested by colony PCR, using primers KAL-1 and KAL-2. The clones containing the predicted *KlALG3 andKlALG61* genes were sequenced and open reading frames identified.

[0177] Primers for construction of an *alg3::NAT*^{R} deletion allele, using a PCR overlap method (Davidson et al, 2002), were designed and the resulting deletion allele was transformed into two *K. lactis* strains and NAT-resistant colonies selected. These colonies were screened by PCR and transformants were obtained in which the *ALG3* ORF was replaced with the *och1::NAT*^{R} mutant allele.

EXAMPLE 2

Generation of an alg3/och1 mutant strain expressing an α-1,2-Mannosidase, GnT1 and GnTII for production of a human-like glycoprotein.

[0178] The 1215 bp open reading frame of the *P. pastoris OCH1* gene as well as 2685 bp upstream and 1175 bp downstream was amplified by PCR (B. K. Choi et al., submitted to *Proc. Natl. Acad. Sci. USA* 2002; see also WO 02/00879; each of which is incorporated herein by reference), cloned into the pCR2.1-TOPO vector

(Invitrogen) and designated pBK9. To create an *och1* knockout strain containing multiple auxotrophic markers, 100 μg of pJN329, a plasmid containing an *och1::URA3* mutant allele flanked with *Sfi*I restriction sites was digested with *Sfi*I and used to transform *P. pastoris* strain JC308 (Cereghino et al. *Gene* 263 (2001) 159-169) by electroporation. Following incubation on defined medium lacking uracil for 10 days at room temperature, 1000 colonies were picked and re-streaked. URA^+ clones that were unable to grow at 37°C, but grew at room temperature, were subjected to colony PCR to test for the correct integration of the *och1::URA3* mutant allele. One clone that exhibited the expected PCR pattern was designated YJN153. The Kringle 3 domain of human plasminogen (K3) was used as a model protein. A Neo^R marked plasmid containing the K3 gene was transformed into strain YJN153 and a resulting strain, expressing K3, was named BK64-1 (B. K. Choi et al, submitted to *Proc. Natl. Acad. Sci. USA* 2002).

[0179] Plasmid pPB103, containing the *Kluyveromyces lactis MNN2-2* gene, encoding a Golgi UDP-N-acetylglucosamine transporter was constructed by cloning a blunt *Bgl*II-*Hind*III fragment from vector pDL02 (Abeijon et al. (1996) *Proc. Natl. Acad. Sci. U.S.A.* 93:5963-5968) into *Bgl*II and *Bam*HI digested and blunt ended pBLADE-SX containing the *P. pastoris ADE1* gene (Cereghino et al. (2001) *Gene* 263:159-169). This plasmid was linearized with *Eco*NI and transformed into strain BK64-1 by electroporation and one strain confirmed to contain the *MNN2-2* by PCR analysis was named PBP1.

[0180] A library of mannosidase constructs was generated, comprising in-frame fusions of the leader domains of several type I or type II membrane proteins from *S. cerevisiae* and *P. pastoris* fused with the catalytic domains of several α-1,2-mannosidase genes from human, mouse, fly, worm and yeast sources (see, e.g., WO02/00879, incorporated herein by reference). This library was created in a *P. pastoris HIS4* integration vector and screened by linearizing with *Sal*I, transforming by electroporation into strain PBP1, and analyzing the glycans released from the K3 reporter protein. One active construct chosen was a chimera of the 988-1296 nucleotides (C-terminus) of the yeast *SEC12* gene fused with a N-terminal deletion of the mouse α-1,2-mannosidase IA (MmMannIA) gene, which

was missing the 187 nucleotides. A *P. pastoris* strain expressing this construct was named PBP2.

[0181] A library of GnTI constructs was generated, comprising in-frame fusions of the same leader library with the catalytic domains of GnTI genes from human, worm, frog and fly sources (WO 02/00879). This library was created in a *P. pastoris ARG4* integration vector and screened by linearizing with *Aat*II, transforming by electroporation into strain PBP2, and analyzing the glycans released from K3. One active construct chosen was a chimera of the first 120 bp of the *S. cerevisiae MNN9* gene fused to a deletion of the human GnTI gene, which was missing the first 154 bp. A *P. pastoris* strain expressing this construct was named PBP3.

[0182] Subsequently, a *P. pastoris alg3::KAN*^{R} deletion construct was generated as described above. Approximately 5μg of the resulting PCR product was transformed into strain PBP3 and colonies were selected on YPD medium containing 200μg/ml G418. One strain out of 20 screened by PCR was confirmed to contain the correct integration of the *alg3::KAN*^{R} mutant allele and lack the wild-type allele. This strain was named RDP27.

[0183] Finally, a library of GnTII constructs was generated, which was comprised of in-frame fusions of the leader library with the catalytic domains of GnTII genes from human and rat sources (WO 02/00879). This library was created in a *P. pastoris* integration vector containing the NST^R gene conferring resistance to the drug nourseothricin. The library plasmids were linearized with *Eco*RI, transformed into strain RDP27 by electroporation, and the resulting strains were screened by analysis of the released glycans from purified K3.

Materials

[0184] MOPS, sodium cacodylate, manganese chloride, UDP-galactose and CMP-N-acetylneuraminic acid were from Sigma. TFA was from Aldrich. Recombinant rat α2,6-sialyltransferase from *Spodoptera frugiperda* and β1,4-galactosyltransferase from bovine milk were from Calbiochem. Protein N-glycosidase F, mannosidases, and oligosaccharides were from Glyko (San Rafael, CA). DEAE ToyoPearl resin was from TosoHaas. Metal chelating "HisBind"

resin was from Novagen (Madison, WI). 96-well lysate-clearing plates were from Promega (Madison, WI). Protein-binding 96-well plates were from Millipore (Bedford, MA). Salts and buffering agents were from Sigma (St. Louis, MO). MALDI matrices were from Aldrich (Milwaukee, WI).

Protein Purification

[0185] Kringle 3 was purified using a 96-well format on a Beckman BioMek 2000 sample-handling robot (Beckman/Coulter Ranch Cucamonga, CA). Kringle 3 was purified from expression media using a C-terminal hexa-histidine tag. The robotic purification is an adaptation of the protocol provided by Novagen for their HisBind resin. Briefly, a 150uL (μL) settled volume of resin is poured into the wells of a 96-well lysate-binding plate, washed with 3 volumes of water and charged with 5 volumes of 50mM NiSO4 and washed with 3 volumes of binding buffer (5mM imidazole, 0.5M NaCl, 20mM Tris-HCL pH7.9). The protein expression media is diluted 3:2, media/PBS (60mM PO4, 16mM KCl, 822mM NaCl pH7.4) and loaded onto the columns. After draining, the columns are washed with 10 volumes of binding buffer and 6 volumes of wash buffer (30mM imidazole, 0.5M NaCl, 20mM Tris-HCl pH7.9) and the protein is eluted with 6 volumes of elution buffer (1M imidazole, 0.5M NaCl, 20mM Tris-HCl pH7.9). The eluted glycoproteins are evaporated to dryness by lyophilyzation.

Release of N-linked Glycans

[0186] The glycans are released and separated from the glycoproteins by a modification of a previously reported method (Papac, et al. A. J. S. (1998) Glycobiology 8, 445-454). The wells of a 96-well MultiScreen IP (Immobilon-P membrane) plate (Millipore) are wetted with 100uL of methanol, washed with 3X150uL of water and 50uL of RCM buffer (8M urea, 360mM Tris, 3.2mM EDTA pH8.6), draining with gentle vacuum after each addition. The dried protein samples are dissolved in 30uL of RCM buffer and transferred to the wells containing 10uL of RCM buffer. The wells are drained and washed twice with RCM buffer. The proteins are reduced by addition of 60uL of 0.1M DTT in RCM buffer for 1hr at 37oC. The wells are washed three times with 300uL of water and

carboxymethylated by addition of 60uL of 0.1M iodoacetic acid for 30min in the dark at room temperature. The wells are again washed three times with water and the membranes blocked by the addition of 100uL of 1% PVP 360 in water for 1hr at room temperature. The wells are drained and washed three times with 300uL of water and deglycosylated by the addition of 30uL of 10mM NH4HCO3 pH 8.3 containing one milliunit of N-glycanase (Glyko). After 16 hours at 37oC, the solution containing the glycans was removed by centrifugation and evaporated to dryness.

Matrix Assisted Laser Desorption Ionization Time of Flight Mass Spectrometry

[0187] Molecular weights of the glycans were determined using a Voyager DE PRO linear MALDI-TOF (Applied Biosciences) mass spectrometer using delayed extraction. The dried glycans from each well were dissolved in 15uL of water and 0.5uL spotted on stainless steel sample plates and mixed with 0.5uL of S-DHB matrix (9mg/mL of dihydroxybenzoic acid, 1mg/mL of 5-methoxysalicilic acid in 1:1 water/acetonitrile 0.1% TFA) and allowed to dry.

[0188] Ions were generated by irradiation with a pulsed nitrogen laser (337nm) with a 4ns pulse time. The instrument was operated in the delayed extraction mode with a 125ns delay and an accelerating voltage of 20kV. The grid voltage was 93.00%, guide wire voltage was 0.10%, the internal pressure was less than 5 X 10-7 torr, and the low mass gate was 875Da. Spectra were generated from the sum of 100-200 laser pulses and acquired with a 2 GHz digitizer. Man5 oligosaccharide was used as an external molecular weight standard. All spectra were generated with the instrument in the positive ion mode. The estimated mass accuracy of the spectra was 0.5%.

Materials:

[0189] MOPS, sodium cacodylate, manganese chloride, UDP-galactose and CMP-N-acetylneuraminic acid were from Sigma, Saint Louis, MO. Trifluroacetic acid (TFA) was from Sigma/Aldrich, Saint Louis, MO. Recombinant rat alpha-2,6-

sialyltransferase from Spodoptera frugiperda and beta-1,4-galactosyltransferase from bovine milk were from Calbiochem, San Diego, CA.

β-N-acetylhexosaminidase Digestion

[0190] The glycans were released and separated from the glycoproteins by a modification of a previously reported method (Papac, et al. A. J. S. (1998) *Glycobiology* 8, 445-454). After the proteins were reduced and carboxymethylated, and the membranes blocked, the wells were washed three time with water. The protein was deglycosylated by the addition of 30 µl of 10 mM NH_4HCO_3 pH 8.3 containing one milliunit of N-glycanase (Glyko, Novato, CA). After 16 hr at 37°C, the solution containing the glycans was removed by centrifugation and evaporated to dryness. The glycans were then dried in SC210A speed vac (Thermo Savant, Halbrook, NY). The dried glycans were put in 50 mM NH_4Ac pH 5.0 at 37°C overnight and 1mU of hexos (Glyko, Novato, CA) was added.

Galactosyltransferase Reaction

[0191] Approximately 2mg of protein (r-K3:hPg [PBP6-5]) was purified by nickel-affinity chromatography, extensively dialyzed against 0.1% TFA, and lyophilized to dryness. The protein was redissolved in 150µL of 50mM MOPS, 20mM MnCl2, pH7.4. After addition of 32.5µg (533nmol) of UDP-galactose and 4mU of β 1,4-galactosyltransferase, the sample was incubated at 37° C for 18 hours. The samples were then dialyzed against 0.1% TFA for analysis by MALDI-TOF mass spectrometry.

[0192] The spectrum of the protein reacted with galactosyltransferase showed an increase in mass consistent with the addition of two galactose moieties when compared with the spectrum of a similar protein sample incubated without enzyme. Protein samples were next reduced, carboxymethylated and deglycosylated with PNGase F. The recovered N-glycans were analyzed by MALDI-TOF mass spectrometry. The mass of the predominant glycan from the galactosyltransferase reacted protein was greater than that of the control glycan by a mass consistent with the addition of two galactose moieties (325.4 Da).

Sialyltransferase Reaction

[0193] After resuspending the (galactosyltransferase reacted) proteins in 10μL of 50mM sodium cacodylate buffer pH6.0, 300μg (488nmol) of CMP-N-acetylneuraminic acid (CMP-NANA) dissolved in 15μL of the same buffer, and 5μL (2mU) of recombinant α-2,6 sialyltransferase were added. After incubation at 37°C for 15 hours, an additional 200μg of CMP-NANA and 1mU of sialyltransferase were added. The protein samples were incubated for an additional 8 hours and then dialyzed and analyzed by MALDI-TOF-MS as above.

[0194] The spectrum of the glycoprotein reacted with sialyltransferase showed an increase in mass when compared with that of the starting material (the protein after galactosyltransferase reaction). The N-glycans were released and analyzed as above. The increase in mass of the two ion-adducts of the predominant glycan was consistent with the addition of two sialic acid residues (580 and 583Da).

EXAMPLE 3
Identification, cloning and deletion of the *ALG9 and ALG 12* genes in *P.pastoris*

[0195] Similar to Example 1, the ALG9p and ALG12 sequences, respectively from *S. cerevisiae, Drosophila melanogaster, Homo sapiens, etc.*, is aligned and regions of high homology are used to design degenerate primers. These primers are employed in a PCR reaction on genomic DNA from the *P. pastoris*. The resulting initial PCR product is subcloned, sequenced and used to probe a Southern blot of genomic DNA from *P. pastoris* with high stringency (Sambrook et al., 1989). Hybridization is observed. This Southern blot is used to map the genomic loci. Genomic fragments are cloned by digesting genomic DNA and ligating those fragments in the appropriate size-range into pUC19 to create a *P. pastoris* subgenomic library. This subgenomic library is transformed into *E. coli* and several hundred clones tested by colony PCR, using primers designed based on the sequence of the initial PCR product. The clones containing the predicted genes are sequenced and open reading frames identified. Primers for construction of an *alg9::NAT*^{R} deletion allele, using a PCR overlap method (Davidson et al., 2002), are designed. The resulting deletion allele is transformed into two *P.pastoris*

strains and NAT resistant colonies are selected. These colonies are screened by PCR and transformants obtained in which the *ALG9* ORF is replaced with the *och1::NAT*^{R} mutant allele. See generally, Cipollo et al. *Glycobiology* 2002 (12)11:749-762; Chantret et al. *J. Biol. Chem.* Jul. 12, 2002 (277)28:25815-25822; Cipollo et al. J. Biol. Chem. Feb. 11, 2000 (275)6:4267-4277; Burda et al. *Proc. Natl. Acad. Sci. U.S.A.* July 1996 (93):7160-7165; Karaoglu et al. *Biochemistry* 2001, 40, 12193-12206; Grimme et al. *J. Biol. Chem.* July 20, 2001 (276)29:27731-27739; Verostek et al. *J. Biol. Chem.* June 5, 1993 (268)16:12095-12103; Huffaker et al. *Proc. Natl. Acad. Sci. U.S.A.* Dec. 1983 (80):7466-7470.

EXAMPLE 4
Identification, cloning and expression of Alpha 1,2-3 Mannosidase From Xanthomonas Manihotis

[0196] The alpha 1,2-3 Mannosidase from *Xanthomonas Manihotis* has two activities: an alpha-1,2 and an alpha-1,3 mannosidase. The methods of the invention may also use two independent mannosidases having these activities, which may be similarly identified and cloned from a selected organism of interest.

[0197] As described by Landry et al., alpha-mannosidases can be purified from *Xanthomonas sp.*, such as *Xanthomonas manihotis*. *X. manihotis* can be purchased from the American Type Culture Collection (ATCC catalog number 49764) (*Xanthomonas axonopodis* Starr and Garces pathovar *manihotis* deposited as *Xanthomonas manihotis* (Arthaud-Berthet) Starr). Enzymes are purified from crude cell-extracts as previously described (Wong-Madden, S.T. and Landry, D. (1995) Purification and characterization of novel glycosidases from the bacterial genus Xanthomonas; and Landry, D. US Patent US 6,300,113 B1 Isolation and composition of novel Glycosidases). After purification of the mannosidase, one of several methods are used to obtain peptide sequence tags (see, e.g., W. Quadroni M et al. (2000). A method for the chemical generation of N-terminal peptide sequence tags for rapid protein identification. *Anal Chem* (2000) Mar 1;72(5):1006-14; Wilkins MR et al. Rapid protein identification using N-terminal "sequence tag" and amino acid analysis. Biochem Biophys Res Commun. (1996)

Apr 25;221(3):609-13; and Tsugita A. (1987) Developments in protein microsequencing. *Adv Biophys* (1987) 23:81-113).

[0198] Sequence tags generated using a method above are then used to generate sets of degenerate primers using methods well-known to the skilled worker. Degenerate primers are used to prime DNA amplification in polymerase chain reactions (e.g., using Taq polymerase kits according to manufacturers' instructions) to amplify DNA fragments. The amplified DNA fragments are used as probes to isolate DNA molecules comprising the gene encoding a desired mannosidase, e.g., using standard Southern DNA hybridization techniques to identify and isolate (clone) genomic pieces encoding the enzyme of interest. The genomic DNA molecules are sequenced and putative open reading frames and coding sequences are identified. A suitable expression construct encoding for the glycosidase of interest can then be generated using methods described herein and well-known in the art.

[0199] Nucleic acid fragments comprising sequences encoding alpha 1,2-3 mannosidase activity (or catalytically active fragments thereof) are cloned into appropriate expression vectors for expression, and preferably targeted expression, of these activities in an appropriate host cell according to the methods set forth herein.

EXAMPLE 5

Identification, cloning and expression of the *ALG6* gene in *P.pastoris*

[0200] Similar to Example 1, the ALG6p sequences from *S. cerevisiae, Drosophila melanogaster, Homo sapiens etc.,* are aligned and regions of high homology are used to design degenerate primers. These primers are employed in a PCR reaction on genomic DNA from the *P. pastoris*. The resulting initial PCR product is subcloned, sequenced and used to probe a Southern blot of genomic DNA from *P. pastoris* with high stringency (Sambrook et al, 1989). Hybridization is observed. This Southern blot is used to map the genomic loci. Genomic fragments are cloned by digesting genomic DNA and ligating those fragments in the appropriate size-range into pUC19 to create a *P. pastoris* subgenomic library. This subgenomic library is transformed into *E. coli* and several hundred clones are

tested by colony PCR, using primers designed based on the sequence of the initial PCR product. The clones containing the predicted genes are sequenced and open reading frames identified. Primers for construction of an *alg6::NAT*^{R} deletion allele, using a PCR overlap method (Davidson et al, 2002), are designed and the resulting deletion allele is transformed into two *P. pastoris* strains and NAT resistant colonies selected. These colonies are screened by PCR and transformants are obtained in which the *ALG6* ORF is replaced with the *och1::NAT*^{R} mutant allele. See, e.g., Imbach et al. *Proc. Natl. Acad. Sci. U.S.A.* June 1999 (96)6982-6987.

[0201] Nucleic acid fragments comprising sequences encoding Alg6p (or catalytically active fragments thereof) are cloned into appropriate expression vectors for expression, and preferably targeted expression, of these activities in an appropriate host cell according to the methods set forth herein. The cloned *ALG6* gene can be brought under the control of any suitable promoter to achieve overexpression. Even expression of the gene under the control of its own promoter is possible. Expression from multicopy plasmids will generate high levels of expression ("overexpression").

EXAMPLE 6
Cloning and Expression Of GnT III To Produce Bisecting GlcNAcs Which Boost Antibody Functionality

A. Background

[0202] The addition of an N-acetylglucosamine to the $GlcNAc_2Man_3GlcNAc_2$ structure by N-acetylglucosaminyltransferases III yields a so-called bisected N-glycan (see **Figure 3**). This structure has been implicated in greater antibody-dependent cellular cytotoxicity (ADCC) (Umana et al. 1999).

[0203] A host cell such as a yeast strain capable of producing glycoproteins with bisected N-glycans is engineered according to the invention, by introducing into the host cell a GnTIII activity. Preferably, the host cell is transformed with a nucleic acid that encodes GnTIII (e.g., a mammalian such as the murine GnT III shown in **Fig. 32**) or a domain thereof having enzymatic activity, optionally fused

to a heterologous cell signal targeting peptide (e.g., using the libraries and associated methods of the invention.)

[0204] IgGs consist of two heavy-chains (V_H, C_H1, C_H2 and C_H3 in Figure 30), interconnected in the hinge region through three disulfide bridges, and two light chains (V_L, C_L in **Figure 30**). The light chains (domains V_L and C_L) are linked by another disulfide bridge to the C_H1 portion of the heavy chain and together with the C_H1 and V_H fragment make up the so-called Fab region. Antigens bind to the terminal portion of the Fab region. The Fc region of IgGs consists of the C_H3, the C_H2 and the hinge region and is responsible for the exertion of so-called effector functions (see below).

[0205] The primary function of antibodies is binding to an antigen. However, unless binding to the antigen directly inactivates the antigen (such as in the case of bacterial toxins), mere binding is meaningless unless so-called effector-functions are triggered. Antibodies of the IgG subclass exert two major effector-functions: the activation of the complement system and induction of phagocytosis. The complement system consists of a complex group of serum proteins involved in controlling inflammatory events, in the activation of phagocytes and in the lytical destruction of cell membranes. Complement activation starts with binding of the C1 complex to the Fc portion of two IgGs in close proximity. C1 consists of one molecule, C1q, and two molecules, C1r and C1s. Phagocytosis is initiated through an interaction between the IgG's Fc fragment and Fc-gamma-receptors (FcγRI, II and III in **Figure 30**). Fc receptors are primarily expressed on the surface of effector cells of the immune system, in particular macrophages, monocytes, myeloid cells and dendritic cells.

[0206] The C_H2 portion harbors a conserved N-glycosylation site at asparagine 297 (Asp297). The Asp297 N-glycans are highly heterogeneous and are known to affect Fc receptor binding and complement activation. Only a minority (i.e., about 15-20%) of IgGs bears a disialylated, and 3-10% have a monosialylated N-glycan (reviewed in Jefferis, R., Glycosylation of human IgG Antibodies. BioPharm, 2001). Interestingly, the minimal N-glycan structure shown to be necessary for fully functional antibodies capable of complement activation and Fc receptor binding is a pentasacharide with terminal N-acetylglucosamine residues

($GlcNAc_2Man_3$) (reviewed in Jefferis, R., Glycosylation of human IgG Antibodies. BioPharm, 2001). Antibodies with less than a $GlcNAc_2Man_3$ N-glycan or no N-glycosylation at Asp297 might still be able to bind an antigen but most likely will not activate the crucial downstream events such as phagocytosis and complement activation. In addition, antibodies with fungal-type N-glycans attached to Asp297 will in all likelihood solicit an immune-response in a mammalian organism which will render that antibody useless as a therapeutic glycoprotein.

B. Cloning And Expression Of GnTIII

The DNA fragment encoding part of the mouse GnTIII protein lacking the TM domain is PCR amplified from murine (or other mammalian) genomic DNA using forward 5'-TCCTGGCGCGCCTTCCCGAGAGAACTGGCCTCCCTC-3' and 5'-AATTAATTAACCCTAGCCCTCCGCTGTATCCAACTTG-3' reversed primers. Those primers include AscI and PacI restriction sites that will be used for cloning into the vector suitable for the fusion with leader library.
The nucleic acid and amino acid sequence of murine GnTIII is shown in **Fig. 32**.

C. Cloning of immunoglobulin encoding sequences

[0207] Protocols for the cloning of the variable regions of antibodies, including primer sequences, have been published previously. Sources of antibodies and encoding genes can be, among others, *in vitro* immunized human B cells (see, e.g., Borreback, C.A. et al. (1988) *Proc. Natl. Acad. Sci. USA* 85, 3995-3999), peripheral blood lymphocytes or single human B cells (see, e.g., Lagerkvist, A.C. et al. (1995) *Biotechniques* 18, 862-869; and Terness, P. et al. (1997) *Hum. Immunol.* 56, 17-27) and transgenic mice containing human immunoglobulin loci, allowing the creation of hybridoma cell-lines.

[0208] Using standard recombinant DNA techniques, antibody-encoding nucleic acid sequences can be cloned. Sources for the genetic information encoding immunoglobulins of interest are typically total RNA preparations from cells of interest, such as blood lymphocytes or hybridoma cell lines. For example, by employing a PCR based protocol with specific primers, variable regions can be cloned via reverse transcription initiated from a sequence-specific primer

hybridizing to the IgG C_H1 domain site and a second primer encoding amino acids 111-118 of the murine kappa constant region. The V_H and V_K encodingcDNAs will then be amplified as previously published (see, e.g., Graziano, R.F. et al. (1995) *J Immunol.* 155(10): p. 4996-5002; Welschof, M. et al. (1995) *J. Immunol. Methods* 179, 203-214; and Orlandi, R. et al. (1988) *Proc. Natl. Acad. Sci. USA* 86: 3833). Cloning procedures for whole immunoglobulins (heavy and light chains have also been published (see, e.g., Buckel, P. et al. (1987) *Gene* 51:13-19; Recinos A 3rd et al. (1994) *Gene* 149: 385-386; (1995) *Gene* Jun 9;158(2):311-2; and Recinos A 3rd et al. (1994) *Gene* Nov 18;149(2):385-6). Additional protocols for the cloning and generation of antibody fragment and antibody expression constructs have been described in Antibody Engineering, R. Kontermann and S. Dübel (2001), Editors, Springer Verlag: Berlin Heidelberg New York.

[0209] Fungal expression plasmids encoding heavy and light chain of immunoglobulins have been described (see, e.g., Abdel-Salam, H.A. et al. (2001) *Appl. Microbiol. Biotechnol.* 56: 157-164; and Ogunjimi, A.A. et al. (1999) *Biotechnology Letters* 21: 561-567). One can thus generate expression plasmids harboring the constant regions of immunoglobulins. To facilitate the cloning of variable regions into these expression vectors, suitable restriction sites can be placed in close proximity to the termini of the variable regions. The constant regions can be constructed in such a way that the variable regions can be easily in-frame fused to them by a simple restriction-digest / ligation experiment. **Figure 31** shows a schematic overview of such an expression construct, designed in a very modular way, allowing easy exchange of promoters, transcriptional terminators, integration targeting domains and even selection markers.

[0210] As shown in **Figure 31**, V_L as well as V_H domains of choice can be easily cloned in-frame with C_L and the C_H regions, respectively. Initial integration is targeted to the P. pastoris AOX locus (or homologous locus in another fungal cell) and the methanol-inducible AOX promoter will drive expression. Alternatively, any other desired constitutive or inducible promoter cassette may be used. Thus, if desired, the 5'AOX and 3'AOX regions as well as transcriptional terminator (TT) fragments can be easily replaced with different TT, promoter and integration targeting domains to optimize expression. Initially the alpha-factor secretion

signal with the standard KEX protease site is employed to facilitate secretion of heavy and light chains. The properties of the expression vector may be further refined using standard techniques.

[0211] An Ig expression vector such as the one described above is introduced into a host cell of the invention that expresses GnTIII, preferably in the Golgi apparatus of the host cell. The Ig molecules expressed in such a host cell comprise N-glycans having bisecting GlcNAcs.

EXAMPLE 7
Cloning and expression of GnT-IV (UDP-GlcNAc:alpha-1,3-D -mannoside beta-1,4-N-Acetylglucosaminyltransferase IV) and GnT-V (beta 1-6-N-acetylglucosaminyltransferase)

[0212] GnTIV-encoding cDNAs were isolated from bovine and human cells (Minowa,M.T. et al. (1998) *J. Biol. Chem.* 273 (19), 11556-11562; and Yoshida,A. et al. (1999) *Glycobiology* 9 (3), 303-310. The DNA fragments encoding full length and a part of the human GnT-IV protein (**Figure 33**) lacking the TM domain are PCR amplified from the cDNA library using forward 5'-AATGAGATGAGGCTCCGCAATGGAACTG-3', 5'-CTGATTGCTTATCAACGAGAATTCCTTG-3', and reverse 5'-TGTTGGTTTCTCAGATGATCAGTTGGTG-3'primers, respectively. The resulting PCR products are cloned and sequenced.

[0213] Similarly, genes encoding GnT-V protein have been isolated from several mammalian species, including mouse. (See, e.g., Alvarez, K. et al. *Glycobiology* 12 (7), 389-394 (2002)). The DNA fragments encoding full length and a part of the mouse GnT-V protein (**Figure 34**) lacking the TM domain are PCR amplified from the cDNA library using forward 5'-AGAGAGAGATGGCTTTCTTTTCTCCCTGG-3', 5'-AAATCAAGTGGATGAAGGACATGTGGC-3', and reverse 5'-AGCGATGCTATAGGCAGTCTTTGCAGAG-3'primers, respectively. The resulting PCR products are cloned and sequenced.

[0214] Nucleic acid fragments comprising sequences encoding GnT IV or V (or catalytically active fragments thereof) are cloned into appropriate expression vectors for expression, and preferably targeted expression, of these activities in an appropriate host cell according to the methods set forth herein.

[0215] REFERENCES

Aebi, M., J. Gassenhuber, et al. (1996). "Cloning and characterization of the ALG3 gene of Saccharomyces cerevisiae." Glycobiology **6**(4): 439-444.

Altmann, F., E. Staudacher, et al. (1999). "Insect cells as hosts for the expression of recombinant glycoproteins." Glycoconjugate Journal **16**(2): 109-123.

Andersen, D. C. and C. F. Goochee (1994). "The effect of cell-culture conditions on the oligosaccharide structures of secreted glycoproteins." Current Opinion in Biotechnology **5**: 546-549.

Bardor, M., L. Faye, et al. (1999). "Analysis of the N-glycosylation of recombinant glycoproteins produced in transgenic plants." Trends in Plant Science **4**(9): 376-380.

Bretthauer, R. K. and F. J. Castellino (1999). "Glycosylation of Pichia pastoris-derived proteins." Biotechnology and Applied Biochemistry **30**: 193-200.

Burda, P. and M. Aebi (1999). "The dolichol pathway of N-linked glycosylation." Biochimica Et Biophysica Acta-General Subjects **1426**(2): 239-257.

Chiba, Y., M. Suzuki, et al. (1998). "Production of human compatible high mannose-type (Man(5)GlcNAc(2)) sugar chains in Saccharomyces cerevisiae." Journal of Biological Chemistry **273**(41): 26298-26304.

Cole, E. S., E. Higgins, et al. (1994). "Glycosylation Patterns of Human Proteins Expressed in Transgenic Goat Milk." Journal of Cellular Biochemistry: 265-265.

Davies et al. Biotechnol Bioeng. 2001 Aug 20;74(4):288-294. (Expression of GnTIII in a Recombinant Anti-CD20 CHO Production Cell Line: Expression of Antibodies with Altered Glycoforms Leads to an Increase in ADCC Through Higher Affinity for FcgRIII).

Dente, L., U. Ruther, et al. (1988). "Expression of Human Alpha-1-Acid Glycoprotein Genes in Cultured-Cells and in Transgenic Mice." Genes & Development **2**(2): 259-266.

Huffaker, T. C. and P. W. Robbins (1983). "Yeast Mutants Deficient in Protein Glycosylation." Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences **80**(24): 7466-7470.

Jarvis, D. L., Z. S. Kawar, et al. (1998). "Engineering N-glycosylation pathways in the baculovirus-insect cell system." Current Opinion in Biotechnology **9**(5): 528-533.

Kimura, T., N. Kitamoto, et al. (1997). "A novel yeast gene, RHK1, is involved in the synthesis of the cell wall receptor for the HM-1 killer toxin that inhibits beta-1,3-glucan synthesis." Molecular & General Genetics **254**(2): 139-147.

Kimura, T., T. Komiyama, et al. (1999). "N-glycosylation is involved in the sensitivity of Saccharomyces cerevisiae to HM-1 killer toxin secreted from Hansenula mrakii IFO 0895." Applied Microbiology and Biotechnology **51**(2): 176-184.

Malissard, M., S. Zeng, et al. (2000). "Expression of functional soluble forms of human beta-1,4- galactosyltransferase I, alpha-2,6-sialyltransferase, and alpha-1,3-fucosyltransferase VI in the methylotrophic yeast Pichia pastoris." Biochemical and Biophysical Research Communications **267**(1): 169-173.

Maras, M. and R. Contreras (1994). Methods of Modifying Carbohydrate Moieties. United States, Alko Group Ltd., Helsinki, Finland.

Maras, M., A. De Bruyn, et al. (1999). "In vivo synthesis of complex N-glycans by expression of human N-acetylglucosaminyltransferase I in the filamentous fungus Trichoderma reesei." Febs Letters **452**(3): 365-370.

Maras, M., X. Saelens, et al. (1997). "In vitro conversion of the carbohydrate moiety of fungal glycoproteins to mammalian-type oligosaccharides - Evidence for N-acetylglucosaminyltransferase-I-accepting glycans from Trichoderma reesei." European Journal of Biochemistry **249**(3): 701-707.

Martinet, W., M. Maras, et al. (1998). "Modification of the protein glycosylation pathway in the methylotrophic yeast Pichia pastoris." Biotechnology Letters **20**(12): 1171-1177.

McGarvey, P. B., J. Hammond, et al. (1995). "Expression of the Rabies Virus Glycoprotein in Transgenic Tomatoes." Bio-Technology **13**(13): 1484-1487.

Moens, S. and J. Vanderleyden (1997). "Glycoproteins in prokaryotes." Archives of Microbiology **168**(3): 169-175.

Nakanishishindo, Y., K. Nakayama, et al. (1993). "Structure of the N-Linked Oligosaccharides That Show the Complete Loss of Alpha-1,6-Polymannose Outer Chain From Och1, Och1 Mnn1, and Och1 Mnn1 Alg3 Mutants of Saccharomyces- Cerevisiae." Journal of Biological Chemistry **268**(35): 26338-26345.

Raju, T. S., J. B. Briggs, et al. (2000). "Species-specific variation in glycosylation of IgG: evidence for the species-specific sialylation and branch-specific galactosylation and importance for engineering recombinant glycoprotein therapeutics." Glycobiology **10**(5): 477-486.

Sharma, C. B., R. Knauer, et al. (2001). "Biosynthesis of lipid-linked oligosaccharides in yeast: the ALG3 gene encodes the DoI-P-Man : Man(5)GlcNAc(2)-PP-DoI mannosyltransferase." Biological Chemistry **382**(2): 321-328.

Staub, J. M., B. Garcia, et al. (2000). "High-yield production of a human therapeutic protein in tobacco chloroplasts." Nature Biotechnology **18**(3): 333-338.

Takeuchi, M. (1997). "Trial for molecular breeding of yeast for the production of glycoprotein therapeutics." Trends in Glycoscience and Glycotechnology **9**: S29-S35.

Umana et al., Nat Biotechnol. 1999a Feb (17)176-180. (Engineered glycoforms of an antineuroblastoma IgG1 with optimized antibodydependent cellular cytotoxic activity)

Umana et al., Biotechnol Bioeng. 1999b Dec 5;65(5):542-549. (Regulated Overexpression of glycosyltransferase).

Verostek, M. F., P. H. Atkinson, et al. (1993). "Glycoprotein-Biosynthesis in the Alg3 Saccharomyces-Cerevisiae Mutant .1. Role of Glucose in the Initial Glycosylation of Invertase in the Endoplasmic-Reticulum." Journal of Biological Chemistry **268**(16): 12095-12103.

Verostek, M. F., P. H. Atkinson, et al. (1993). "Glycoprotein-Biosynthesis in the Alg3 Saccharomyces-Cerevisiae Mutant .2. Structure of Novel Man6-

10glcnac2 Processing Intermediates On Secreted Invertase." Journal of Biological Chemistry **268**(16): 12104-12115.

Weikert, S., D. Papac, et al. (1999). "Engineering Chinese hamster ovary cells to maximize sialic acid content of recombinant glycoproteins." Nature Biotechnology **17**(11): 1116-1121.

Werner, R. G., W. Noe, et al. (1998). "Appropriate mammalian expression systems for biopharmaceuticals." Arzneimittel-Forschung-Drug Research **48**(8): 870-880.

Yang, M. and M. Butler (2000). "Effects of ammonia on CHO cell growth, erythropoietin production, and glycosylation." Biotechnology and Bioengineering **68**(4): 370-380.Zufferey, R., R. Knauer, et al. (1995). "Stt3, a Highly Conserved Protein Required for Yeast Oligosaccharyl Transferase-Activity in-Vivo." EMBO Journal **14**(20): 4949-4960.